



INVESTOR IN PEOPLE

The Patent Office
 Concept House
 Cardiff Road
 Newport
 South Wales
 NP10 8QQ

CD 06 JAN 2004

WIPO

PCT

PRIORITY DOCUMENT

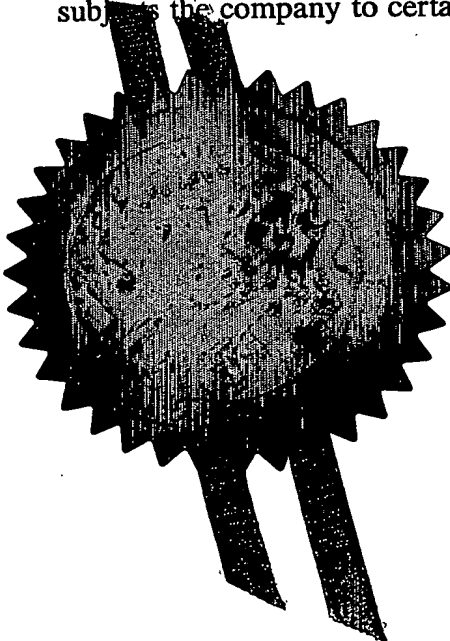
SUBMITTED OR TRANSMITTED IN
 COMPLIANCE WITH RULE 17.1(a) OR (b)

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.



Signed

Dated

19 December 2003

Patents Form 1/77

Patents Act 1977
(Rule 16)

The
Patent
Office

THE PATENT
C
21 NOV 2002
LONDON

1/77

The Patent Office
Cardiff Road
Newport
Gwent NP9 1RH

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form.)

1. Your reference 42.1.78522

2. Patent application number 0227238.3

3. Full name, address and postcode of the
or of each applicant (underline all surnames)

DiaGenic AS
Østensjøveien 15B
0661 Oslo
Norway

08511552001

Patents ADP number (if you know it)

If the applicant is a corporate body, give
country/state of incorporation

4. Title of the invention Product and Method

5. Name of your agent (if you have one) Frank B. Dehn & Co.

"Address for service" in the United Kingdom
to which all correspondence should be sent
(including the postcode)

179 Queen Victoria Street
London
EC4V 4EL

Patents ADP number (if you know it)

166001

6. If you are declaring priority from one or more
earlier patent applications, give the country
and the date of filing of the or of each of these
earlier applications and (if you know it) the or
each application number

Country

Priority application number
(if you know it)

Date of filing
(day / month / year)

7. If this application is divided or otherwise
derived from an earlier UK application,
give the number and the filing date of
the earlier application

Number of earlier application

Date of filing
(day / month / year)

8. Is a statement of inventorship and of right
to grant of a patent required in support of
this request? (Answer 'Yes' if:

- a) any applicant named in part 3 is not an inventor, or
- b) there is an inventor who is not named as an
applicant, or
- c) any named applicant is a corporate body.
See note (d))

yes

6 Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	0
Description	115
Claim(s)	0
Abstract	0
Drawing(s)	3 + 3 RM

10. If you are also filing any of the following, state how many against each item.

Priority documents

Translations of priority documents	0
Statement of inventorship and right to grant of a patent (Patents Form 7/77)	0
Request for preliminary examination and search (Patents Form 9/77)	0
Request for substantive examination (Patents Form 10/77)	0
Any other documents (please specify)	0

11.

I/We request the grant of a patent on the basis of this application.

Frank B. Deane

Signature

Date 21 November 2002

12. Name and daytime telephone number of person to contact in the United Kingdom

Elizabeth Jones
020 7206 0600

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

- If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.
- Write your answers in capital letters using black ink or you may type them.
- If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s) of the form. Any continuation sheet should be attached to this form.
- If you have answered 'Yes', Patents Form 7/77 will need to be filed.
- Once you have filled in the form you must remember to sign and date it.
- For details of the fee and ways to pay please contact the Patent Office.

78522.618

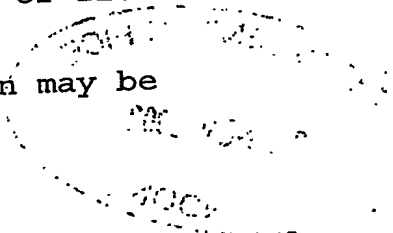
Product and Method

5 The present invention relates to oligonucleotide
probes, for use in assessing gene transcript levels in a
cell, which may be used in analytical techniques,
particularly diagnostic techniques. Conveniently the
probes are provided in kit form. Different sets of
10 probes may be used in techniques to prepare gene
expression patterns and identify, diagnose or monitor
different states, such as diseases, conditions or stages
thereof. Also provided are methods of identifying
suitable probes and their use in methods of the
15 invention.

 The identification of quick and easy methods of
sample analysis for, for example, diagnostic
applications, remains the goal of many researchers. End
users seek methods which are cost effective, produce
20 statistically significant results and which may be
implemented routinely without the need for highly
skilled individuals.

 The analysis of gene expression within cells has
been used to provide information on the state of those
25 cells and importantly the state of the individual from
which the cells are derived. The relative expression of
various genes in a cell has been identified as
reflecting a particular state within a body. For
example, cancer cells are known to exhibit altered
30 expression of various proteins and the transcripts or
the expressed proteins may therefore be used as markers
of that disease state.

 Thus biopsy tissue may be analysed for the presence
of these markers and cells originating from the site of
35 the disease may be identified in other tissues or fluids
of the body by the presence of the markers.
Furthermore, products of the altered expression may be



released into the blood stream and these products may be analysed. In addition cells which have contacted disease cells may be affected by their direct contact with those cells resulting in altered gene expression and their expression or products of expression may be similarly analysed.

However, there are some limitations with these methods. For example, the use of specific tumour markers for identifying cancer suffers from a variety of defects, such as lack of specificity or sensitivity, association of the marker with disease states besides the specific type of cancer, and difficulty of detection in asymptomatic individuals.

In addition to the analysis of one or two marker transcripts or proteins, more recently, gene expression patterns have been analysed. Most of the work involving large-scale gene expression analysis with implications in disease diagnosis has involved clinical samples originating from diseased tissues or cells. For example, several recent publications, which demonstrate that gene expression data can be used to distinguish between similar cancer types, have used clinical samples from diseased tissues or cells (Alon et al. 1999, PNAS, 96, p6745-6750; Golub et al. 1999, Science, 286, p531-537; Alizadeh et al, 2000, Nature, 403, p503-511; Bittner et al., 2000, Nature, 406, p536-540).

However, these methods have relied on analysis of a sample containing diseased cells or products of those cells or cells which have been contacted by disease cells. Analysis of such samples relies on knowledge of the presence of a disease and its location, which may be difficult in asymptomatic patients. Furthermore, samples can not always be taken from the disease site, e.g. in diseases of the brain.

In a finding of great significance, the present inventors identified the previously untapped potential of all cells within a body to provide information

relating to the state of the organism from which the cells were derived. WO98/49342 describes the analysis of the gene expression of cells distant from the site of disease, e.g. peripheral blood collected distant from a cancer site.

This finding is based on the premise that the different parts of an organism's body exist in dynamic interaction with each other. When a disease affects one part of the body, other parts of the body are also affected. The interaction results from a wide spectrum of biochemical signals that are released from the diseased area, affecting other areas in the body. Although, the nature of the biochemical and physiological changes induced by the released signals can vary in the different body parts, the changes can be measured at the level of gene expression and used for diagnostic purposes.

The physiological state of a cell in an organism is determined by the pattern with which genes are expressed in it. The pattern depends upon the internal and external biological stimuli to which said cell is exposed, and any change either in the extent or in the nature of these stimuli can lead to a change in the pattern with which the different genes are expressed in the cell. There is a growing understanding that by analysing the systemic changes in gene expression patterns in cells in biological samples, it is possible to provide information on the type and nature of the biological stimuli that are acting on them. Thus, for example, by monitoring the expression of a large number of genes in cells in a test sample, it is possible to determine whether their genes are expressed with a pattern characteristic for a particular disease, condition or stage thereof. Measuring changes in gene activities in cells, e.g. from tissue or body fluids is therefore emerging as a powerful tool for disease diagnosis.

Such methods have various advantages. Often, obtaining clinical samples from certain areas in the body that is diseased can be difficult and may involve undesirable invasions in the body, for example biopsy is often used to obtain samples for cancer. In some cases, such as in Alzheimer's disease the diseased brain specimen can only be obtained post-mortem. Furthermore, the tissue specimens which are obtained are often heterogeneous and may contain a mixture of both diseased and non-diseased cells, making the analysis of generated gene expression data both complex and difficult.

It has been suggested that a pool of tumour tissues that appear to be pathogenetically homogeneous with respect to morphological appearances of the tumour may well be highly heterogeneous at the molecular level. (Alizadeh, 2000, supra), and in fact might contain tumours representing essentially different diseases (Alizadeh, 2000, supra; Golub, 1999, supra). For the purpose of identifying a disease, condition, or a stage thereof, any method that does not require clinical samples to originate directly from diseased tissues or cells is highly desirable since clinical samples representing a homogeneous mixture of cell types can be obtained from an easily accessible region in the body.

In an extension of the above described work, we now describe probes and sets of probes derived from cells which are not disease cells and which have not contacted disease cells, which correspond to genes which exhibit altered expression in normal versus disease individuals, for use in methods of identifying, diagnosing or monitoring certain conditions, particularly diseases or stages thereof.

Thus the invention provides a set of oligonucleotide probes which correspond to genes in a cell whose expression is affected in a pattern characteristic of a particular disease, condition or stage thereof, wherein said genes are systemically

affected by said disease, condition or stage thereof. Preferably said genes are metabolic or house-keeping genes and preferably are moderately or highly expressed. Preferably the genes are moderately or highly expressed in the cells of the sample but not in cells from disease cells or in cells having contacted such cells.

Such probes, particularly when isolated from cells distant to the site of disease, do not rely on the development of disease to clinically recognizable levels and allow detection of a disease or condition or stage thereof very early after the onset of said disease or condition, even years before other subjective or objective symptoms appear.

As used herein "systemically" affected genes refers to genes whose expression is affected in the body without direct contact with a disease cell or disease site and the cells under investigation are not disease cells.

"Contact" as referred to herein refers to cells coming into close proximity with one another such that the direct effect of one cell on the other may be observed, e.g. an immune response, wherein these responses are not mediated by secondary molecules released from the first cell over a large distance to affect the second cell. Preferably contact refers to physical contact, or contact that is as close as is sterically possible, conveniently, cells which contact one another are found in the same unit volume, for example within 1cm^3 .

A "disease cell" is a cell manifesting phenotypic changes and is present at the disease site at some time during its life-span, e.g. a tumour cell at the tumour site or which has disseminated from the tumour, or a brain cell in the case of brain disorders such as Alzheimer's disease.

"Metabolic" or "house-keeping" genes refer to those genes responsible for expressing products involved in

cell division and maintenance, e.g. non-immune function related genes.

"Moderately or highly" expressed genes refers to those present in resting cells in a copy number of more than 30-100 copies/cell (assuming an average 3×10^5 mRNA molecules in a cell.

Specific probes having the above described properties are provided herein.

Thus in one aspect, the present invention provides a set of oligonucleotide probes, wherein said set comprises at least 10 oligonucleotides selected from:

an oligonucleotide as described in Table 1, or derived from a sequence described in Table 1, or an oligonucleotide with a complementary sequence, or a functionally equivalent oligonucleotide.

The invention also provides one or more oligonucleotide probes, wherein each oligonucleotide probe is selected from the oligonucleotides listed in Table 1, or derived from a sequence described in Table 1, or a complementary sequence thereof. The use of such probes in products and methods of the invention, form further aspects of the invention.

As referred to herein an "oligonucleotide" is a nucleic acid molecule having at least 6 monomers in the polymeric structure, ie. nucleotides or modified forms thereof. The nucleic acid molecule may be DNA, RNA or PNA (peptide nucleic acid) or hybrids thereof or modified versions thereof, e.g. chemically modified forms, e.g. LNA (Locked Nucleic Acid), by methylation or made up of modified or non-natural bases during synthesis, providing they retain their ability to bind to complementary sequences. Such oligonucleotides are used in accordance with the invention to probe target sequences and are thus referred to herein also as oligonucleotide probes or simply as probes.

An "oligonucleotide derived from a sequence described in Table 1" (or any other table) refers to a

part of a sequence disclosed in that Table (e.g. Table 1-4), which satisfies the requirements of the oligonucleotide probes as described herein, e.g. in length and function. Preferably said parts have the size described hereinafter.

Preferably the oligonucleotide probes forming said set are at least 15 bases in length to allow binding of target molecules. Especially preferably said oligonucleotide probes are from 20 to 200 bases in length, e.g. from 30 to 150 bases, preferably 50-100 bases in length.

As referred to herein the term "complementary sequences" refers to sequences with consecutive complementary bases (ie. T:A, G:C) and which complementary sequences are therefore able to bind to one another through their complementarity.

Reference to "10 oligonucleotides" refers to 10 different oligonucleotides. Whilst a Table 1 oligonucleotide, a Table 1 derived oligonucleotide and their functional equivalents are considered different oligonucleotides, complementary oligonucleotides are not considered different. Preferably however, the at least 10 oligonucleotides are 10 different Table 1 oligonucleotides (or Table 1 derived oligonucleotides or their functional equivalents). Thus said 10 different oligonucleotides are preferably able to bind to 10 different transcripts.

Preferably said oligonucleotides are as described in Table 1 or are derived from a sequence described in Table 1. Especially preferably said oligonucleotides are as described in Table 2 or Table 4 or are derived from a sequence described in either of those tables. Especially preferably the oligonucleotide (or the oligonucleotide derived therefrom) has a high occurrence as defined in Table 3, especially preferably >40%, e.g. >80 or >90, e.g. 100%.

A "set" as described refers to a collection of

unique oligonucleotide probes (ie. having a distinct sequence) and preferably consists of less than 1000 oligonucleotide probes, especially less than 500 probes, e.g. preferably from 10 to 500, e.g. 10 to 100, 200 or 300, especially preferably 20 to 100, e.g. 30 to 100 probes. In some cases less than 10 probes may be used, e.g. from 2 to 9 probes, e.g. 5 to 9 probes.

It will be appreciated that increasing the number of probes will prevent the possibility of poor analysis, e.g. misdiagnosis by comparison to other diseases which could similarly alter the expression of the particular genes in question. Other oligonucleotide probes not described herein may also be present, particularly if they aid the ultimate use of the set of oligonucleotide probes. However, preferably said set consists only of said Table 1 oligonucleotides, Table 1 derived oligonucleotides, complementary sequences or functionally equivalent oligonucleotides, or a sub-set thereof (e.g. of the size as described above). Especially preferably said set consists only of said Table 1 oligonucleotides, Table 1 derived oligonucleotides, or complementary sequences thereof, or a sub-set thereof.

Multiple copies of each unique oligonucleotide probe, e.g. 10 or more copies, may be present in each set, but constitute only a single probe.

A set of oligonucleotide probes, which may preferably be immobilized on a solid support or have means for such immobilization, comprises the at least 10 oligonucleotide probes selected from those described hereinbefore. Especially preferably said probes are selected from those having high occurrence as described in Table 3 and as mentioned above. As mentioned above, these 10 probes must be unique and have different sequences. Having said this however, two separate probes may be used which recognize the same gene but reflect different splicing events. However

oligonucleotide probes which are complementary to, and bind to distinct genes are preferred.

As described herein a "functionally equivalent" oligonucleotide to those described in Table 1 or derived therefrom refers to an oligonucleotide which is capable of identifying the same gene as an oligonucleotide of Table 1 or derived therefrom, ie. it can bind to the same mRNA molecule (or DNA) transcribed from a gene (target nucleic acid molecule) as the Table 1 oligonucleotide or the Table 1 derived oligonucleotide (or its complementary sequence). Preferably said functionally equivalent oligonucleotide is capable of recognizing, ie. binding to the same splicing product as a Table 1 oligonucleotide or a Table 1 derived oligonucleotide. Preferably said mRNA molecule is the full length mRNA molecule which corresponds to the Table 1 oligonucleotide or the Table 1 derived oligonucleotide.

As referred to herein "capable of binding" or "binding" refers to the ability to hybridize under conditions described hereinafter.

Alternatively expressed, functionally equivalent oligonucleotides (or complementary sequences) have sequence identity or will hybridize, as described hereinafter, to a region of the target molecule to which molecule a Table 1 oligonucleotide or a Table 1 derived oligonucleotide or a complementary oligonucleotide binds. Preferably, functionally equivalent oligonucleotides hybridize to one of the mRNA sequences which corresponds to a Table 1 oligonucleotide or a Table 1-derived-oligonucleotide under the conditions described hereinafter or has sequence identity to a part of one of the mRNA sequences which corresponds to a Table 1 oligonucleotide or a Table 1 derived oligonucleotide. A "part" in this context refers to a stretch of at least 5, e.g. at least 10 or 20 bases, such as from 5 to 100, e.g. 10 to 50 or 15 to 30 bases.

In a particularly preferred aspect, the functionally equivalent oligonucleotide binds to all or a part of the region of a target nucleic acid molecule (mRNA or cDNA) to which the Table 1 oligonucleotide or
5 Table 1 derived oligonucleotide binds. A "target" nucleic acid molecule is the gene transcript or related product e.g. mRNA, or cDNA, or amplified product thereof. Said "region" of said target molecule to which said Table 1 oligonucleotide or Table 1 derived
10 oligonucleotide binds is the stretch over which complementarity exists. At its largest this region is the whole length of the Table 1 oligonucleotide or Table 1 derived oligonucleotide, but may be shorter if the entire Table 1 sequence or Table 1 derived
15 oligonucleotide is not complementary to a region of the target sequence.

Preferably said part of said region of said target molecule is a stretch of at least 5, e.g. at least 10 or 20 bases, such as from 5 to 100, e.g. 10 to 50 or 15 to
20 30 bases. This may for example be achieved by said functionally equivalent oligonucleotide having several identical bases to the bases of the Table 1 oligonucleotide or the Table 1 derived oligonucleotide. These bases may be identical over consecutive stretches,
25 e.g. in a part of the functionally equivalent oligonucleotide, or may be present non-consecutively, but provide sufficient complementarity to allow binding to the target sequence.

Thus in a preferred feature, said functionally
30 equivalent oligonucleotide hybridizes under conditions of high stringency to a Table 1 oligonucleotide or a Table 1 derived oligonucleotide or the complementary sequence thereof. Alternatively expressed, said functionally equivalent oligonucleotide exhibits high
35 sequence identity to all or part of a Table 1 oligonucleotide. Preferably said functionally equivalent oligonucleotide has at least 70% sequence

identity, preferably at least 80%, e.g. at least 90, 95, 98 or 99%, to all of a Table 1 oligonucleotide or a part thereof. As used in this context, a "part" refers to a stretch of at least 5, e.g. at least 10 or 20 bases, such as from 5 to 100, e.g. 10 to 50 or 15 to 30 bases, in said Table 1 oligonucleotide. Especially preferably when sequence identity to only a part of said Table 1 oligonucleotide is present, the sequence identity is high, e.g. at least 80% as described above.

Functionally equivalent oligonucleotides which satisfy the above stated functional requirements include those which are derived from the Table 1 oligonucleotides and also those which have been modified by single or multiple nucleotide base (or equivalent) substitution, addition and/or deletion, but which nonetheless retain functional activity, e.g. bind to the same target molecule as the Table 1 oligonucleotide or the Table 1 derived oligonucleotide from which they are further derived or modified. Preferably said modification is of from 1 to 50, e.g. from 10 to 30, preferably from 1 to 5 bases. Especially preferably only minor modifications are present, e.g. variations in less than 10 bases, e.g. less than 5 base changes.

Within the meaning of "addition" equivalents are included oligonucleotides containing additional sequences which are complementary to the consecutive stretch of bases on the target molecule to which the Table 1 oligonucleotide or the Table 1 derived oligonucleotide binds. Alternatively the addition may comprise a different, unrelated sequence, which may for example confer a further property, e.g. to provide a means for immobilization such as a linker to bind the oligonucleotide probe to a solid support.

Particularly preferred are naturally occurring equivalents such as biological variants, e.g. allelic, geographical or allotypic variants, e.g. oligonucleotides which correspond to a genetic variant,

for example as present in a different species.

Functional equivalents include oligonucleotides with modified bases, e.g. using non-naturally occurring bases. Such derivatives may be prepared during synthesis or by post production modification.

"Hybridizing" sequences which bind under conditions of low stringency are those which bind under non-stringent conditions (for example, 6x SSC/50% formamide at room temperature) and remain bound when washed under conditions of low stringency (2 X SSC, room temperature, more preferably 2 X SSC, 42°C). Hybridizing under high stringency refers to the above conditions in which washing is performed at 2 X SSC, 65°C (where SSC = 0.15M NaCl, 0.015M sodium citrate, pH 7.2).

"Sequence identity" as referred to herein refers to the value obtained when assessed using ClustalW (Thompson et al., 1994, Nucl. Acids Res., 22, p4673-4680) with the following parameters:

Pairwise alignment parameters - Method: accurate,
Matrix: IUB, Gap open penalty: 15.00, Gap extension penalty: 6.66;
Multiple alignment parameters - Matrix: IUB, Gap open penalty: 15.00, % identity for delay: 30, Negative matrix: no, Gap extension penalty: 6.66, DNA transitions weighting: 0.5.

Sequence identity at a particular base is intended to include identical bases which have simply been derivatized.

The invention also extends to polypeptides encoded by the mRNA sequence to which a Table 1 oligonucleotide or a Table 1 derived oligonucleotide binds. The invention further extends to antibodies which bind to any of said polypeptides.

As described above, conveniently said set of oligonucleotide probes may be immobilized on one or more solid supports. Single or preferably multiple copies of each unique probe are attached to said solid supports,

e.g. 10 or more, e.g. at least 100 copies of each unique probe are present.

One or more unique oligonucleotide probes may be associated with separate solid supports which together form a set of probes immobilized on multiple solid support, e.g. one or more unique probes may be immobilized on multiple beads, membranes, filters, biochips etc. which together form a set of probes, which together form modules of the kit described hereinafter.

The solid support of the different modules are conveniently physically associated although the signals associated with each probe (generated as described hereinafter) must be separately determinable.

Alternatively, the probes may be immobilized on discrete portions of the same solid support, e.g. each unique oligonucleotide probe, e.g. in multiple copies, may be immobilized to a distinct and discrete portion or region of a filter or membrane, e.g. to generate an array.

A combination of such techniques may also be used, e.g. several solid supports may be used which each immobilize several unique probes.

The expression "solid support" shall mean any solid material able to bind oligonucleotides by hydrophobic, ionic or covalent bridges.

"Immobilization" as used herein refers to reversible or irreversible association of the probes to said solid support by virtue of such binding. If reversible, the probes remain associated with the solid support for a time sufficient for methods of the invention to be carried out.

Numerous solid supports suitable as immobilizing moieties according to the invention, are well known in the art and widely described in the literature and generally speaking, the solid support may be any of the well-known supports or matrices which are currently widely used or proposed for immobilization, separation

etc. in chemical or biochemical procedures. Such materials include, but are not limited to, any synthetic organic polymer such as polystyrene, polyvinylchloride, polyethylene; or nitrocellulose and cellulose acetate; or tosyl activated surfaces; or glass or nylon or any surface carrying a group suited for covalent coupling of nucleic acids. The immobilizing moieties may take the form of particles, sheets, gels, filters, membranes, microfibre strips, tubes or plates, fibres or capillaries, made for example of a polymeric material e.g. agarose, cellulose, alginate, teflon, latex, polystyrene or magnetic beads. Solid supports allowing the presentation of an array, preferably in a single dimension are preferred, e.g. sheets, filters, membranes, plates or biochips.

Attachment of the nucleic acid molecules to the solid support may be performed directly or indirectly. For example if a filter is used, attachment may be performed by UV-induced crosslinking. Alternatively, attachment may be performed indirectly by the use of an attachment moiety carried on the oligonucleotide probes and/or solid support. Thus for example, a pair of affinity binding partners may be used, such as avidin, streptavidin or biotin, DNA or DNA binding protein (e.g. either the lac I repressor protein or the lac operator sequence to which it binds), antibodies (which may be mono- or polyclonal), antibody fragments or the epitopes or haptens of antibodies. In these cases, one partner of the binding pair is attached to (or is inherently part of) the solid support and the other partner is attached to (or is inherently part of) the nucleic acid molecules.

As used herein an "affinity binding pair" refers to two components which recognize and bind to one another specifically (ie. in preference to binding to other molecules). Such binding pairs when bound together form a complex.

Attachment of appropriate functional groups to the solid support may be performed by methods well known in the art, which include for example, attachment through hydroxyl, carboxyl, aldehyde or amino groups which may be provided by treating the solid support to provide suitable surface coatings. Solid supports presenting appropriate moieties for attachment of the binding partner may be produced by routine methods known in the art.

Attachment of appropriate functional groups to the oligonucleotide probes of the invention may be performed by ligation or introduced during synthesis or amplification, for example using primers carrying an appropriate moiety, such as biotin or a particular sequence for capture.

Conveniently, the set of probes described hereinbefore is provided in kit form.

Thus viewed from a further aspect the present invention provides a kit comprising a set of oligonucleotide probes as described hereinbefore immobilized on one or more solid supports.

Preferably, said probes are immobilized on a single solid support and each unique probe is attached to different region of said solid support. However, when attached to multiple solid supports, said multiple solid supports form the modules which make up the kit. Especially preferably said solid support is a sheet, filter, membrane, plate or biochip.

Optionally the kit may also contain information relating to the signals generated by normal or diseased samples (as discussed in more detail hereinafter in relation to the use of the kits), standardizing materials, e.g. mRNA or cDNA from normal and/or diseased samples for comparative purposes, labels for incorporation into cDNA, adapters for introducing nucleic acid sequences for amplification purposes, primers for amplification and/or appropriate enzymes,

5 buffers and solutions. Optionally said kit may also contain a package insert describing how the method of the invention should be performed, optionally providing standard graphs, data or software for interpretation of results obtained when performing the invention.

The use of such kits to prepare a standard diagnostic gene transcript pattern as described hereinafter forms a further aspect of the invention.

10 The set of probes as described herein have various uses. Principally however they are used to assess the gene expression state of a test cell to provide information relating to the organism from which said cell is derived. Thus the probes are useful in diagnosing, identifying or monitoring a disease or
15 condition or stage thereof in an organism.

Thus in a further aspect the invention provides the use of a set of oligonucleotide probes or a kit as described hereinbefore to determine the gene expression pattern of a cell which pattern reflects the level of
20 gene expression of genes to which said oligonucleotide probes bind, comprising at least the steps of:

- a) isolating mRNA from said cell, which may optionally be reverse transcribed to cDNA;
- b) hybridizing the mRNA or cDNA of step (a) to a
25 set of oligonucleotide probes or a kit as defined herein; and
- c) assessing the amount of mRNA or cDNA hybridizing to each of said probes to produce said pattern.

30 The mRNA and cDNA as referred to in this method, and the methods hereinafter, encompass derivatives or copies of said molecules, e.g., copies of such molecules such as those produced by amplification or the preparation of complementary strands, but which retain the identity of the mRNA sequence, ie. would hybridize
35 to the direct transcript (or its complementary sequence) by virtue of precise complementarity, or sequence identity, over at least a region of said molecule. It

will be appreciated that complementarity will not exist over the entire region where techniques have been used which may truncate the transcript or introduce new sequences, e.g. by primer amplification. For convenience, said mRNA or cDNA is preferably amplified prior to step b). As with the oligonucleotides described herein said molecules may be modified, e.g. by using non-natural bases during synthesis providing complementarity remains. Such molecules may also carry additional moieties such as signalling or immobilizing means.

The various steps involved in the method of preparing such a pattern are described in more detail hereinafter.

As used herein "gene expression" refers to transcription of a particular gene to produce a specific mRNA product (ie. a particular splicing product). The level of gene expression may be determined by assessing the level of transcribed mRNA molecules or cDNA molecules reverse transcribed from the mRNA molecules or products derived from those molecules, e.g. by amplification.

The "pattern" created by this technique refers to information which, for example, may be represented in tabular or graphical form and conveys information about the signal associated with two or more oligonucleotides. Preferably said pattern is expressed as an array of numbers relating to the expression level associated with each probe.

Preferably, said pattern is established using the following linear model:

$$y = Xb + f$$

Equation 1

wherein, X is the matrix of gene expression data and y is the response variable, b is the regression coefficient vector and f the estimated residual vector. Although many different methods can be used to establish the relationship provided in equation 1, especially

preferably the partial Least Squares Regression (PLSR) method is used for establishing the relationship in equation 1.

5 The probes are thus used to generate a pattern
which reflects the gene expression of a cell at the time
of its isolation. The pattern of expression is
characteristic of the circumstances under which that
cells finds itself and depends on the influences to
10 which the cell has been exposed. Thus, a characteristic
gene transcript pattern standard or fingerprint
(standard probe pattern) for cells from an individual
with a particular disease or condition may be prepared
and used for comparison to transcript patterns of test
cells. This has clear applications in diagnosing,
15 monitoring or identifying whether an organism is
suffering from a particular disease, condition or stage
thereof.

 The standard pattern is prepared by determining the
extent of binding of total mRNA (or cDNA or related
20 product), from cells from a sample of one or more
organisms with the disease or condition or stage
thereof, to the probes. This reflects the level of
transcripts which are present which correspond to each
unique probe. The amount of nucleic acid material which
25 binds to the different probes is assessed and this
information together forms the gene transcript pattern
standard of that disease or condition or stage thereof.
Each such standard pattern is characteristic of the
disease, condition or stage thereof.

30 In a further aspect therefore, the present
invention provides a method of preparing a standard gene
transcript pattern characteristic of a disease or
condition or stage thereof in an organism comprising at
least the steps of:

35 a) isolating mRNA from the cells of a sample of one
or more organisms having the disease or condition or
stage thereof, which may optionally be reverse

transcribed to cDNA;

b) hybridizing the mRNA or cDNA of step (a) to a set of oligonucleotides or a kit as described hereinbefore specific for said disease or condition or stage thereof in an organism and sample thereof corresponding to the organism and sample thereof under investigation; and

c) assessing the amount of mRNA or cDNA hybridizing to each of said probes to produce a characteristic pattern reflecting the level of gene expression of genes to which said oligonucleotides bind, in the sample with the disease, condition or stage thereof.

For convenience, said oligonucleotides are preferably immobilized on one or more solid supports.

The standard pattern for a great number of diseases or conditions and different stages thereof using particular probes may be accumulated in databases and be made available to laboratories on request.

"Disease" samples and organisms as referred to herein refer to organisms (or samples from the same) with an underlying pathological disturbance relative to a normal organism (or sample), in a symptomatic or asymptomatic organism, which may result, for example, from infection or an acquired or congenital genetic imperfection. Such organisms are known to have, or which exhibit, the disease or condition or stage thereof under study.

A "condition" refers to a state of the mind or body of an organism which has not occurred through disease, e.g. the presence of an agent in the body such as a toxin, drug or pollutant; or pregnancy.

"Stages" thereof refer to different stages of the disease or condition which may or may not exhibit particular physiological or metabolic changes, but do exhibit changes at the genetic level which may be detected as altered gene expression. It will be appreciated that during the course of a disease or

condition the expression of different transcripts may vary. Thus at different stages, altered expression may not be exhibited for particular transcripts compared to "normal" samples. However, combining information from
5 several transcripts which exhibit altered expression at one or more stages through the course of the disease or condition can be used to provide a characteristic pattern which is indicative of a particular stage of the disease or condition. Thus for example different stages
10 in cancer, e.g. pre-stage I, stage I, stage II, II or IV can be identified.

"Normal" as used herein refers to organisms or samples which are used for comparative purposes. Preferably, these are "normal" in the sense that they do
15 not exhibit any indication of, or are not believed to have, any disease or condition that would affect gene expression, particularly in respect of the disease for which they are to be used as the normal standard. However, it will be appreciated that different stages of
20 a disease or condition may be compared and in such cases, the "normal" sample may correspond to the earlier stage of the disease or condition.

As used herein a "sample" refers to any material obtained from the organism, e.g. human or non-human
25 animal under investigation which contains cells and includes, tissues, body fluid or body waste or in the case of prokaryotic organisms, the organism itself. "Body fluids" include blood, saliva, spinal fluid, semen, lymph. "Body waste" includes urine, expectorated
30 matter (pulmonary patients), faeces etc. "Tissue samples" include tissue obtained by biopsy, by surgical interventions or by other means e.g. placenta. Preferably however, the samples which are examined are
35 from areas of the body not apparently affected by the disease or condition. The cells in such samples are not disease cells, e.g. cancer cells, have not been in contact with such cells and do not originate from the

site of the disease or condition. The "site of disease" is considered to be that area of the body which manifests the disease in a way which may be objectively determined, e.g. a tumour or area of inflammation. Thus
5 for example peripheral blood may be used for the diagnosis of non-haematopoietic cancers, and the blood does not require the presence of malignant or disseminated cells from the cancer in the blood. Similarly in diseases of the brain, in which no diseased
10 cells are found in the blood due to the blood:brain barrier, peripheral blood may still be used in the methods of the invention.

It will however be appreciated that the method of preparing the standard transcription pattern and other
15 methods of the invention are also applicable for use on living parts of eukaryotic organisms such as cell lines and organ cultures and explants.

As used herein, reference to "corresponding" sample etc. refers to cells preferably from the same tissue,
20 body fluid or body waste, but also includes cells from tissue, body fluid or body waste which are sufficiently similar for the purposes of preparing the standard. When used in reference to genes "corresponding" to the probes, this refers to genes which are related by
25 sequence (which may be complementary) to the probes although the probes may reflect different splicing products of expression.

"Assessing" as used herein refers to both quantitative and qualitative assessment which may be
30 determined in absolute or relative terms.

The invention may be put into practice as follows:
To prepare a standard transcript pattern for a particular disease, condition or stage thereof, sample mRNA is extracted from the cells of tissues, body fluid
35 or body waste according to known techniques (see for example Sambrook et. al. (1989), Molecular Cloning : A laboratory manual, 2nd Ed., Cold Spring Harbor

Laboratory Press, Cold Spring Harbor, N.Y.) from a diseased individual or organism.

Owing to the difficulties in working with RNA, the RNA is preferably reverse transcribed at this stage to form first strand cDNA. Cloning of the cDNA or selection from, or using, a cDNA library is not however necessary in this or other methods of the invention. Preferably, the complementary strands of the first strand cDNAs are synthesized, ie. second strand cDNAs, but this will depend on which relative strands are present in the oligonucleotide probes. The RNA may however alternatively be used directly without reverse transcription and may be labelled if so required.

Preferably the cDNA strands are amplified by known amplification techniques such as the polymerase chain reaction (PCR) by the use of appropriate primers. Alternatively, the cDNA strands may be cloned with a vector, used to transform a bacteria such as E. coli which may then be grown to multiply the nucleic acid molecules. When the sequence of the cDNAs are not known, primers may be directed to regions of the nucleic acid molecules which have been introduced. Thus for example, adapters may be ligated to the cDNA molecules and primers directed to these portions for amplification of the cDNA molecules. Alternatively, in the case of eukaryotic samples, advantage may be taken of the polyA tail and cap of the RNA to prepare appropriate primers.

To produce the standard diagnostic gene transcript pattern or fingerprint for a particular disease or condition or stage thereof, the above described oligonucleotide probes are used to probe mRNA or cDNA of the diseased sample to produce a signal for hybridization to each particular oligonucleotide probe species, ie. each unique probe. A standard control gene transcript pattern may also be prepared if desired using mRNA or cDNA from a normal sample. Thus, mRNA or cDNA is brought into contact with the oligonucleotide probe

under appropriate conditions to allow hybridization.

When multiple samples are probed, this may be performed consecutively using the same probes, e.g. on one or more solid supports, ie. on probe kit modules, or
5 by simultaneously hybridizing to corresponding probes, e.g. the modules of a corresponding probe kit.

To identify when hybridization occurs and obtain an indication of the number of transcripts/cDNA molecules which become bound to the oligonucleotide probes, it is
10 necessary to identify a signal produced when the transcripts (or related molecules) hybridize (e.g. by detection of double stranded nucleic acid molecules or detection of the number of molecules which become bound, after removing unbound molecules, e.g. by washing).

15 In order to achieve a signal, either or both components which hybridize (ie. the probe and the transcript) carry or form a signalling means or a part thereof. This "signalling means" is any moiety capable of direct or indirect detection by the generation or
20 presence of a signal. The signal may be any detectable physical characteristic such as conferred by radiation emission, scattering or absorption properties, magnetic properties, or other physical properties such as charge, size or binding properties of existing molecules (e.g.
25 labels) or molecules which may be generated (e.g. gas emission etc.). Techniques are preferred which allow signal amplification, e.g. which produce multiple signal events from a single active binding site, e.g. by the catalytic action of enzymes to produce multiple
30 detectable products.

Conveniently the signalling means may be a label which itself provides a detectable signal. Conveniently this may be achieved by the use of a radioactive or other label which may be incorporated during cDNA
35 production, the preparation of complementary cDNA strands, during amplification of the target mRNA/cDNA or added directly to target nucleic acid molecules.

Appropriate labels are those which directly or indirectly allow detection or measurement of the presence of the transcripts/cDNA. Such labels include for example radiolabels, chemical labels, for example
5 chromophores or fluorophores (e.g. dyes such as fluorescein and rhodamine), or reagents of high electron density such as ferritin, haemocyanin or colloidal gold. Alternatively, the label may be an enzyme, for example peroxidase or alkaline phosphatase, wherein the presence
10 of the enzyme is visualized by its interaction with a suitable entity, for example a substrate. The label may also form part of a signalling pair wherein the other member of the pair is found on, or in close proximity to, the oligonucleotide probe to which the
15 transcript/cDNA binds, for example, a fluorescent compound and a quench fluorescent substrate may be used. A label may also be provided on a different entity, such as an antibody, which recognizes a peptide moiety attached to the transcripts/cDNA, for example attached
20 to a base used during synthesis or amplification.

A signal may be achieved by the introduction of a label before, during or after the hybridization step. Alternatively, the presence of hybridizing transcripts may be identified by other physical properties, such as
25 their absorbance, and in which case the signalling means is the complex itself.

The amount of signal associated with each oligonucleotide probe is then assessed. The assessment may be quantitative or qualitative and may be based on
30 binding of a single transcript species (or related cDNA or other products) to each probe, or binding of multiple transcript species to multiple copies of each unique probe. It will be appreciated that quantitative results will provide further information for the transcript
35 fingerprint of the disease which is compiled. This data may be expressed as absolute values (in the case of macroarrays) or may be determined relative to a

particular standard or reference e.g. a normal control sample.

Furthermore it will be appreciated that the standard diagnostic gene pattern transcript may be prepared using one or more disease samples (and normal samples if used) to perform the hybridization step to obtain patterns not biased towards a particular individual's variations in gene expression.

The use of the probes to prepare standard patterns and the standard diagnostic gene transcript patterns thus produced for the purpose of identification or diagnosis or monitoring of a particular disease or condition or stage thereof in a particular organism forms a further aspect of the invention.

Once a standard diagnostic fingerprint or pattern has been determined for a particular disease or condition using the selected oligonucleotide probes, this information can be used to identify the presence, absence or extent or stage of that disease or condition in a different test organism or individual.

To examine the gene expression pattern of a test sample, a test sample of tissue, body fluid or body waste containing cells, corresponding to the sample used for the preparation of the standard pattern, is obtained from a patient or the organism to be studied. A test gene transcript pattern is then prepared as described hereinbefore as for the standard pattern.

In a further aspect therefore, the present invention provides a method of preparing a test gene transcript pattern comprising at least the steps of:

a) isolating mRNA from the cells of a sample of said test organism, which may optionally be reverse transcribed to cDNA;

b) hybridizing the mRNA or cDNA of step (a) to a set of oligonucleotides or a kit as described hereinbefore specific for a disease or condition or stage thereof in an organism and sample thereof

corresponding to the organism and sample thereof under investigation; and

- 5 c) assessing the amount of mRNA or cDNA hybridizing to each of said probes to produce said pattern reflecting the level of gene expression of genes to which said oligonucleotides bind, in said test sample.

10 This test pattern may then be compared to one or more standard patterns to assess whether the sample contains cells having the disease, condition or stage thereof.

Thus viewed from a further aspect the present invention provides a method of diagnosing or identifying or monitoring a disease or condition or stage thereof in an organism, comprising the steps of:

- 15 a) isolating mRNA from a sample of said organism, which may optionally be reverse transcribed to cDNA;
- 20 b) hybridizing the mRNA or cDNA of step (a) to a set of oligonucleotides or a kit as described hereinbefore specific for said disease or condition or stage thereof in an organism and sample thereof corresponding to the organism and sample thereof under investigation;
- 25 c) assessing the amount of mRNA or cDNA hybridizing to each of said probes to produce a characteristic pattern reflecting the level of gene expression of genes to which said oligonucleotides bind, in said sample; and
- 30 d) comparing said pattern to a standard diagnostic pattern prepared according to the method of the invention using a sample from an organism corresponding to the organism and sample under investigation to determine the presence of said disease or condition or a
- 35 stage thereof in the organism under investigation.

The method up to and including step c) is the

preparation of a test pattern as described above.

As referred to herein, "diagnosis" refers to determination of the presence or existence of a disease or condition or stage thereof in an organism.

5 "Monitoring" refers to establishing the extent of a disease or condition, particularly when an individual is known to be suffering from a disease or condition, for example to monitor the effects of treatment or the development of a disease or condition, e.g. to determine
10 the suitability of a treatment or provide a prognosis.

The presence of the disease or condition or stage thereof may be determined by determining the degree of correlation between the standard and test samples' patterns. This necessarily takes into account the range
15 of values which are obtained for normal and diseased samples. Although this can be established by obtaining standard deviations for several representative samples binding to the probes to develop the standard, it will be appreciated that single samples may be sufficient to
20 generate the standard pattern to identify a disease if the test sample exhibits close enough correlation to that standard. Conveniently, the presence, absence, or extent of a disease or condition or stage thereof in a test sample can be predicted by inserting the data
25 relating to the expression level of informative probes in test sample into the standard diagnostic probe pattern established according to equation 1.

Data generated using the above mentioned methods may be analysed using various techniques from the most
30 basic visual representation (e.g. relating to intensity) to more complex data manipulation to identify underlying patterns which reflect the interrelationship of the level of expression of each gene to which the various probes bind, which may be quantified and expressed
35 mathematically. Conveniently, the raw data thus generated may be manipulated by the data processing and statistical methods described hereinafter, particularly

normalizing and standardizing the data and fitting the data to a classification model to determine whether said test data reflects the pattern of a particular disease, condition or stage thereof.

5 The methods described herein may be used to identify, monitor or diagnose a disease, condition or ailment or its stage or progression, for which the oligonucleotide probes are informative. "Informative" probes as described herein, are those which reflect
10 genes which have altered expression in the diseases or conditions in question, or particular stages thereof. Probes of the invention may not be sufficiently informative for diagnostic purposes when used alone, but
15 are informative when used as one of several probes to provide a characteristic pattern, e.g. in a set as described hereinbefore.

 Preferably said probes correspond to genes which are systemically affected by said disease, condition or stage thereof. Especially preferably said genes are
20 metabolic or house-keeping genes and preferably are moderately or highly expressed. The advantage of using probes directed to moderately or highly expressed genes is that smaller clinical samples are required for generating the necessary gene expression data set, e.g.
25 less than 1ml samples.

 In preferred methods of the invention, the set of probes of the invention are informative for a variety of different diseases, conditions or stages thereof. A sub-set of the probes disclosed herein may be used for
30 diagnosis, identification or monitoring a particular disease, condition or stage thereof.

 Thus the probes may be used to diagnose or identify or monitor any condition, ailment, disease or reaction that leads to the relative increase or decrease in the
35 activity of informative genes of any or all eukaryotic or prokaryotic organisms regardless of whether these changes have been caused by the influence of bacteria,

virus, prions, parasites, fungi, radiation, natural or artificial toxins, drugs or allergens, including mental conditions due to stress, neurosis, psychosis or deteriorations due to the ageing of the organism, and conditions or diseases of unknown cause, providing a sub-set of the probes as described herein are informative for said disease or condition or stage thereof.

Such diseases include those which result in metabolic or physiological changes, such as fever-associated diseases such as influenza or malaria. Other diseases which may be detected include for example yellow fever, sexually transmitted diseases such as gonorrhea, fibromyalgia, candida-related complex, cancer (for example of the stomach, lung, breast, prostate gland, bowel, skin etc), Alzheimer's disease, disease caused by retroviruses such as HIV, senile dementia, multiple sclerosis and Creutzfeldt-Jakob disease to mention a few.

The invention may also be used to identify patients with psychiatric or psychosomatic diseases such as schizophrenia and eating disorders. Of particular importance is the use of this method to detect diseases, conditions, or stages thereof, which are not readily detectable by known diagnostic methods, such as HIV which is generally not detectable using known techniques 1 to 4 months following infection. Conditions which may be identified include for example drug abuse, such as the use of narcotics, alcohol, steroids or performance enhancing drugs.

Preferably said disease to be identified or monitored is a cancer or a degenerative brain disorder (such as Alzheimer's or Parkinson's disease).

In particular, a set of oligonucleotide probes, wherein said set comprises at least 10 oligonucleotides selected from:

an oligonucleotide as described in Table 4 or an

oligonucleotide derived therefrom or an
oligonucleotide with a complementary sequence, or a
functionally equivalent oligonucleotide,

may be used for diagnosis or identification or

5 monitoring the progression of Alzheimer's disease.

Similarly Table 2 probes and Table 2 derived probes and
their functional equivalents may be used to diagnose,
identify or monitor the progression of breast cancer.

Especially preferably the probes used for breast cancer
10 analysis are selected based on their occurrence as set
forth in Table 3 and as described hereinbefore.

The diagnostic method may be used alone as an
alternative to other diagnostic techniques or in

15 addition to such techniques. For example, methods of
the invention may be used as an alternative or additive
diagnostic measure to diagnosis using imaging techniques
such as Magnetic Resonance Image (MRI), ultrasound
imaging, nuclear imaging or X-ray imaging, for example
in the identification and/or diagnosis of tumours.

20 The methods of the invention may be performed on
cells from prokaryotic or eukaryotic organisms which may
be any eukaryotic organisms such as human beings, other
mammals and animals, birds, insects, fish and plants,
and any prokaryotic organism such as a bacteria.

25 Preferred non-human animals on which the methods of
the invention may be conducted include, but are not
limited to mammals, particularly primates, domestic
animals, livestock and laboratory animals. Thus
preferred animals for diagnosis include mice, rats,
30 guinea pigs, cats, dogs, pigs, cows, goats, sheep,
horses. Particularly preferably the disease state or
condition of humans is diagnosed, identified or
monitored.

35 As described above, the sample under study may be
any convenient sample which may be obtained from an
organism. Preferably however, as mentioned above, the
sample is obtained from a site distant to the site of

disease and the cells in such samples are not disease cells, have not been in contact with such cells and do not originate from the site of the disease or condition.

5 It has been found that the cells from such samples show significant and informative variations in the gene expression of a large number of genes. Thus, the same probe (or several probes) may be found to be informative in determinations regarding two or more diseases, conditions or stages thereof by virtue of the particular
10 level of transcripts binding to that probe or the interrelationship of the extent of binding to that probe relative to other probes. As a consequence, it is possible to use a relatively small number of probes for screening for multiple disorders or diseases. This has
15 consequences with regard to the selection of probes, discussed in relation to random identification of probes hereinafter, but also for the use of a single set of probes for more than one diagnosis.

20 Thus, the present invention also provides sets of probes for diagnosing, identifying or monitoring two or more diseases, conditions or stages thereof, wherein at least one of said probes is suitable for said diagnosing, identifying or monitoring at least two of said diseases, conditions or stages thereof, and kits
25 and methods of using the same. Preferably at least 5 probes, e.g. from 5 to 15 probes, are used in at least two diagnoses.

30 Thus, in a further preferred aspect, the present invention provides a method of diagnosis or identification or monitoring as described hereinbefore for the diagnosis, identification or monitoring of two or more diseases, conditions or stages thereof in an organism, wherein said test pattern produced in step c) of the diagnostic method is compared in step d) to at
35 least two standard diagnostic patterns prepared as described previously, wherein each standard diagnostic pattern is a pattern generated for a different disease

or condition or stage thereof.

Whilst in a preferred aspect the methods of assessment concern the development of a gene transcript pattern from a test sample and comparison of the same to a standard pattern, the elevation or depression of expression of certain markers may also be examined by examining the products of expression and the level of those products. Thus a standard pattern in relation to the expressed product may be generated.

In such methods the levels of expression of a set of polypeptides encoded by the gene to which an oligonucleotide of Table 1 or a Table 1 derived oligonucleotide, binds, are analysed.

Various diagnostic methods may be used to assess the amount of polypeptides (or fragments thereof) which are present. The presence or concentration of polypeptides may be examined, for example by the use of a binding partner to said polypeptide (e.g. an antibody), which may be immobilized, to separate said polypeptide from the sample and the amount of polypeptide may then be determined.

"Fragments" of the polypeptides refers to a domain or region of said polypeptide, e.g. an antigenic fragment, which is recognizable as being derived from said polypeptide to allow binding of a specific binding partner. Preferably such a fragment comprises a significant portion of said polypeptide and corresponds to a product of normal post-synthesis processing.

Thus in a further aspect the present invention provides a method of preparing a standard gene transcript pattern characteristic of a disease or condition or stage thereof in an organism comprising at least the steps of:

a) releasing target polypeptides from a sample of one or more organisms having the disease or condition or stage thereof;

b) contacting said target polypeptides with one or

more binding partners, wherein each binding partner is specific to a marker polypeptide (or a fragment thereof) encoded by the gene to which an oligonucleotide of Table 1 (or derived from a sequence described in Table 1) binds, to allow binding of said binding partners to said target polypeptides, wherein said marker polypeptides are specific for said disease or condition thereof in an organism and sample thereof corresponding to the organism and sample thereof under investigation; and

10 c) assessing the target polypeptide binding to said binding partners to produce a characteristic pattern reflecting the level of gene expression of genes which express said marker polypeptides, in the sample with the disease, condition or stage thereof.

15 As used herein "target polypeptides" refer to those polypeptides present in a sample which are to be detected and "marker polypeptides" are polypeptides which are encoded by the genes to which Table 1 oligonucleotides or Table 1 derived oligonucleotides bind. The target and marker polypeptides are identical or at least have areas of high similarity, e.g. epitopic regions to allow recognition and binding of the binding partner.

25 "Release" of the target polypeptides refers to appropriate treatment of a sample to provide the polypeptides in a form accessible for binding of the binding partners, e.g. by lysis of cells where these are present. The samples used in this case need not necessarily comprise cells as the target polypeptides may be released from cells into the surrounding tissue or fluid, and this tissue or fluid may be analysed, e.g. urine or blood. Preferably however the preferred samples as described herein are used. "Binding partners" comprise the separate entities which together make an affinity binding pair as described above, wherein one partner of the binding pair is the target or marker polypeptide and the other partner binds

specifically to that polypeptide, e.g. an antibody.

Various arrangements may be envisaged for detecting the amount of binding pairs which form. In its simplest form, a sandwich type assay e.g. an immunoassay such as
5 an ELISA, may be used in which an antibody specific to the polypeptide and carrying a label (as described elsewhere herein) may be bound to the binding pair (e.g. the first antibody:polypeptide pair) and the amount of label detected.

10 Other methods as described herein may be similarly modified for analysis of the protein product of expression rather than the gene transcript and related nucleic acid molecules.

The methods of generating standard and test
15 patterns and diagnostic techniques rely on the use of informative oligonucleotide probes to generate the gene expression data. In some cases it will be necessary to select these informative probes for a particular method, e.g. to diagnose a particular disease, from a selection
20 of available probes, e.g. the probes described hereinbefore (the Table 1 oligonucleotides, the Table 1 derived oligonucleotides, their complementary sequences and functionally equivalent oligonucleotides). The following methodology describes a convenient method for
25 identifying such informative probes, or more particularly how to select a suitable sub-set of probes from the probes described herein.

Probes for the analysis of a particular disease or condition or stage thereof, may be identified in a
30 number of ways known in the prior art, including by differential expression or by library subtraction (see for example WO98/49342). As described hereinafter, in view of the high information content of most transcripts, as a starting point one may also simply
35 analyse a random sub-set of mRNA or cDNA species and pick the most informative probes from that sub-set. The following method describes the use of immobilized

oligonucleotide probes (e.g. the probes of the invention) to which mRNA (or related molecules) from different samples is bound to identify which probes are the most informative to identify a particular type of sample, e.g. a disease sample.

The immobilized probes can be derived from various unrelated or related organisms; the only requirement is that the immobilized probes should bind specifically to their homologous counterparts in test organisms. Probes can also be derived from commercially available or public databases and immobilized on solid supports or, as mentioned above, they can be randomly picked and isolated from a cDNA library and immobilized on a solid support.

The length of the probes immobilised on the solid support should be long enough to allow for specific binding to the target sequences. The immobilised probes can be in the form of DNA, RNA or their modified products or PNAs (peptide nucleic acids). Preferably, the probes immobilised should bind specifically to their homologous counterparts representing highly and moderately expressed genes in test organisms. Conveniently the probes which are used are the probes described herein.

The gene expression pattern of cells in biological samples can be generated using prior art techniques such as microarray or macroarray as described below or using methods described herein. Several technologies have now been developed for monitoring the expression level of a large number of genes simultaneously in biological samples, such as, high-density-oligoarrays (Lockhart-et al., 1996, Nat. Biotech., 14, p1675-1680), cDNA microarrays (Schena et al, 1995, Science, 270, p467-470) and cDNA macroarrays (Maier E et al., 1994, Nucl. Acids Res., 22, p3423-3424; Bernard et al., 1996, Nucl. Acids Res., 24, p1435-1442).

In high-density oligoarrays and cDNA microarrays,

hundreds and thousands of probe oligonucleotides or cDNAs, are spotted onto glass slides or nylon membranes, or synthesized on biochips. The mRNA isolated from the test and reference samples are labelled by reverse transcription with a red or green fluorescent dye, mixed, and hybridised to the microarray. After washing, the bound fluorescent dyes are detected by a laser, producing two images, one for each dye. The resulting ratio of the red and green spots on the two images provides the information about the changes in expression levels of genes in the test and reference samples.

In cDNA macroarray, different cDNAs are spotted on a solid support, such as nylon membranes in excess in relation to the amount of test mRNA that can hybridise to each spot. mRNA isolated from test samples is radio-labelled by reverse transcription and hybridised to the immobilised probe cDNA. After washing, the signals associated with labels hybridising specifically to immobilised probe cDNA are detected and quantified. The data obtained in macroarray contains information about the relative levels of transcripts present in the test samples. Whilst macroarrays are only suitable to monitor the expression of a limited number of genes, microarrays can be used to monitor the expression of several thousand genes simultaneously and is, therefore, a preferred choice for large-scale gene expression studies.

A macroarray technique for generating the gene expression data set has been used to illustrate the probe identification method described herein. For this purpose, mRNA is isolated from samples of interest and used to prepare labelled target molecules, e.g. mRNA or cDNA as described above. The labelled target molecules are then hybridised to probes immobilised on the solid support. Various solid supports can be used for the purpose, as described previously. Following hybridization, unbound target molecules are removed and

signals from target molecules hybridizing to immobilised probes quantified. If radio labelling is performed, PhosphoImager can be used to generate an image file that can be used to generate a raw data set. Depending on
5 the nature of label chosen for labelling the target molecules, other instruments can also be used, for example, when fluorescence is used for labelling, a FluoroImager can be used to generate an image file from the hybridised target molecules.

10 The raw data corresponding to mean intensity, median intensity, or volume of the signals in each spot can be acquired from the image file using commercially available software for image analysis. However, the acquired data needs to be corrected for background
15 signals and normalized prior to analysis, since, several factors can affect the quality and quantity of the hybridising signals. For example, variations in the quality and quantity of mRNA isolated from sample to sample, subtle variations in the efficiency of labelling
20 target molecules during each reaction, and variations in the amount of unspecific binding between different macroarrays can all contribute to noise in the acquired data set that must be corrected for prior to analysis.

Background correction can be performed in several
25 ways. The lowest pixel intensity within a spot can be used for background subtraction or the mean or median of the line of pixels around the spots' outline can be used for the purpose. One can also define an area representing the background intensity based on the
30 signals generated from negative controls and use the average intensity of this area for background subtraction.

The background corrected data can then be transformed for stabilizing the variance in the data
35 structure and normalized for the differences in probe intensity. Several transformation techniques have been described in the literature and a brief overview can be

found in Cui, Kerr and Churchill

<http://www.jax.org/research/churchill/research/>
expression/Cui-Transform.pdf). Normalization can be

5 performed by dividing the intensity of each spot with
the collective intensity, average intensity or median
intensity of all the spots in a macroarray or a group of
spots in a macroarray in order to obtain the relative
intensity of signals hybridising to immobilised probes
in a macroarray. Several methods have been described
10 for normalizing gene expression data (Richmond and
Somerville, 2000, *Current Opin. Plant Biol.*, 3, p108-
116; Finkelstein et al., 2001; In "Methods of Microarray
Data Analysis. Papers from CAMDA, Eds. Lin & Johnson,
Kluwer Academic, p57-68; Yang et al., 2001, In "Optical
15 Technologies and Informatics", Eds. Bittner, Chen,
Dorsel & Dougherty, *Proceedings of SPIE*, 4266, p141-152;
Dudoit et al, 2000, *J. Am. Stat. Ass.*, 97, p77-87; Alter
et al 2000, *supra*; Newton et al., 2001, *J. Comp. Biol.*,
8, p37-52). Generally, a scaling factor or function is
20 first calculated to correct the intensity effect and
then used for normalising the intensities. The use of
external controls has also been suggested for improved
normalization.

One other major challenge encountered in
25 large-scale gene expression analysis is that of
standardization of data collected from experiments
performed at different times. We have observed that
gene expression data for samples acquired in the same
experiment can be efficiently compared following
30 background correction and normalization. However, the
data from samples acquired in experiments performed at
different times requires further standardization prior
to analysis. This is because subtle differences in
experimental parameters between different experiments,
35 for example, differences in the quality and quantity of
mRNA extracted at different times, differences in time
used for target molecule labelling, hybridization time

or exposure time, can affect the measured values. Also, factors such as the nature of the sequence of transcripts under investigation (their GC content) and their amount in relation to the each other determines how they are affected by subtle variations in the experimental processes. They determine, for example, how efficiently first strand cDNAs, corresponding to a particular transcript, are transcribed and labelled during first strand synthesis, or how efficiently the corresponding labelled target molecules bind to their complementary sequences during hybridization. Failure to properly address and rectify for these influences leads to situations where the differences between the experimental series may overshadow the main information of interest contained in the gene expression data set, i.e. the differences within the combined data from the different experimental series. Figure 1 provides one such example showing a classification based on Principal Component Analysis (PCA) of combined data from two experimental series where the main goal is to distinguish between Alzheimer/non-Alzheimer patients.

PCA (also known as singular value decomposition) is a technique for studying interdependencies and underlying relationships of a set of variables. The data are modelled in terms of a few significant factors or principal components (PC's), plus residuals. The PC's contain the main phenomena and define the systematic variability present in the data, while the residuals represent the variability interpreted as noise. Details on PCA can be found in Jolliffe (1986, Principal Component Analysis, Springer-Verlag, NY), and Jackson (1991, A User's Guide to Principal Components, Wiley, NY). The results of Figure 1 show that two clusters are formed representing the data from two experimental series rather than the Alzheimer/non-Alzheimer differentiation. There were eight samples in common between the two series of

experiments, which ideally should have fallen on top of or in near proximity to each other if appropriately standardized.

We have now found that gene expression data between different experiments can be efficiently standardized by including a subset of samples from one experimental series in the next experimental series and using a direct standardization method (DS), originally described by Wang and Kowalski (Anal. Chem., 1991, 63, p2750 and J. Chemometrics, 1991, 5, p129-145). Although the method of DS is well known in the field of analytical chemistry, it remains undescribed and unused in the field of gene expression data analysis.

In DS, the secondary data representing for example experimental series 2 (secondary measurements, R_2) are corrected to match the data measured on the primary measurements representing data from series 1 (R_1), while the calibration model remains unchanged. In DS, response matrices for both experimental series are related to each other by a transformation matrix F , i.e.

$$R_1 = R_2 F \quad (1)$$

Where F is a square matrix dimensioned gene by gene. From (1), the transformation matrix is calculated as:

$$F = R_2^+ R_1 \quad (2)$$

The transformation matrix F in equation (2) is calculated using a relatively small subset of samples which are measured on both the master primary and the secondary series of data.

Finally, the response of the unknown sample measured on the secondary series $r_{2,un}^T$, is standardized to the response vector $\hat{r}_{1,un}^T$ expected from the primary series

$$\hat{r}_{1,un}^T = r_{21,un}^T \hat{F} \quad (3)$$

From the preceding equation it can be seen that the column i of the transformation matrix contains the multiplication factors for a set of genes measured in the secondary series to obtain the intensity at spot i of the corrected series.

The number of samples that are repeated in the experimental series, R_1 and R_2 , should be equal to their ranks, which in this case is equal to the number of principal components retained for explaining the variation in the R_1 and R_2 . For example, if three principal components are retained for explaining the variation in the data set, a minimum of three samples should be repeated between R_1 and R_2 . The samples that should be repeated between different series should ideally be those that exhibit high leverages in the gene expression pattern. At times, two samples may suffice, while at other times, more than two samples should be ideally be included for good representativity. In some cases, the samples selected can be the same in all the experimental series to be compared (reference samples), while in other cases, representative samples can be selected sequentially by analyzing the expression pattern after each experiment. The selected samples with high leverages are then included in the next experimental series. The results of using Direct Standardization are shown in Figure 1.

Another approach for normalizing and standardizing the gene expression data set is to hybridize each DNA array with target molecules prepared from a test sample and an equal amount of labelled target molecules prepared from representative reference samples. In order to measure the intensity of labelled target molecules hybridizing to the immobilized probes it is necessary that the labelled molecules are prepared from test and reference samples using different labels, for

example, different fluorescent dyes can be used for preparing the labelled material. The labelled molecules prepared from reference samples can be added to the hybridization solution together with the labelled material prepared from test samples. A data file from each array representing the expression pattern of different genes in the test sample and reference samples can then be obtained, normalized and standardized by the direct standardization method as described above. An instant advantage of including the differentially labelled target molecules from reference samples during hybridization is that it enables an efficient comparison of new test samples to the data sets already stored in a database.

Monitoring the expression of a large number of genes in several samples leads to the generation of a large amount of data that is too complex to be easily interpreted. Several unsupervised and supervised multivariate data analysis techniques have already been shown to be useful in extracting meaningful biological information from these large data sets. Cluster analysis is by far the most commonly used technique for gene expression analysis, and has been performed to identify genes that are regulated in a similar manner, and or identifying new/unknown tumour classes using gene expression profiles (Eisen et al., 1998, PNAS, 95, p14863-14868; Alizadeh et al. 2000, supra; Perou et al. 2000, Nature, 406, p747-752; Ross et al, 2000, Nature Genetics, 24(3), p227-235; Herwig et al., 1999, Genome Res., 9, p1093-1105; Tamayo et al, 1999, Science, PNAS, 96, p2907-2912).

In the clustering method, genes are grouped into functional categories (clusters) based on their expression profile, satisfying two criteria: *homogeneity* - the genes in the same cluster are highly similar in expression to each other; and *separation* - genes in different clusters have low similarity in expression to

each other.

Examples of various clustering techniques that have been used for gene expression analysis include hierarchical clustering (Eisen et al., 1998, supra; Alizadeh et al. 2000, supra; Perou et al. 2000, supra; Ross et al, 2000, supra), K-means clustering (Herwig et al., 1999, supra; Tavazoie, 1999, Nature Genetics, 22(3), p281-285), gene shaving (Hastie et al., 2000, Genome Biology, 1(2), research 0003.1-0003.21), block clustering (Tibshirani et al., 1999, Tech repot Univ Stanford.) Plaid model (Lazzeroni, 2002, Stat. Sinica, 12, p61-86), and self-organizing maps (Tamayo et al. 1999, supra). Also, related methods of multivariate statistical analysis, such as those using the singular value decomposition (Alter et al., 2000, PNAS, 97(18), p10101-10106; Ross et al. 2000, supra) or multidimensional scaling can be effective at reducing the dimensions of the objects under study.

However, methods such as cluster analysis and singular value decomposition are purely exploratory and only provide a broad overview of the internal structure present in the data. They are unsupervised approaches in which the available information concerning the nature of the class under investigation is not used in the analysis. Often, the nature of the biological perturbation to which a particular sample has been subjected is known. For example, it is sometimes known whether the sample whose gene expression pattern is being analysed derives from a diseased or healthy individual. In such instances, discriminant analysis can be used for classifying samples into various groups based on their gene expression data.

In such an analysis one builds the classifier by training the data that is capable of discriminating between member and non-members of a given class. The trained classifier can then be used to predict the class of unknown samples. Examples of discrimination methods

that have been described in the literature include Support Vector Machines (Brown et al, 2000, PNAS, 97, p262-267), Nearest Neighbour (Dudoit et al, 2000, supra), Classification trees (Dudoit et al, 2000, supra), Voted classification (Dudoit et al, 2000, supra) Weighted Gene voting (Golub et al. 1999, supra), and Bayesian classification (Keller et al., 2000, Tec report Univ of Washington). Also a technique in which PLS (Partial Least Square) regression analysis is first used to reduce the dimensions in the gene expression data set followed by classification using logistic discriminant analysis and quadratic discriminant analysis (LD and QDA) has recently been described (Nguyen & Rocke, 2002, Bioinformatics, 18, p39-50 and 1216-1226).

A challenge that gene expression data poses to classical discriminatory methods is that the number of genes whose expression are being analysed is very large compared to the number of samples being analysed. However in most cases only a small fraction of these genes are informative in discriminant analysis problems. Moreover, there is a danger that the noise from irrelevant genes can mask or distort the information from the informative genes. Several methods have been suggested in literature to identify and select genes that are informative in microarray studies, for example, t-statistics (Dudoit et al, 2002, J. Am. Stat. Ass., 97, p77-87), analysis of variance (Kerr et al., 2000, PNAS, 98, p8961-8965), Neighbourhood analysis (Golub et al, 1999, supra), Ratio of between groups to within groups sum of squares (Dudoit et al., 2002, supra), Non parametric scoring (Park et al., 2002, Pacific Symposium.... on Biocomputing, p52-63) and Likelihood selection (Keller et al., 2000, supra).

In the methods described herein the gene expression data that has been normalized and standardized is analysed by using Partial Least Squares Regression (PLSR). Although PLSR is primarily a method used for

regression analysis of continuous data (see Appendix A), it can also be utilized as a method for model building and discriminant analysis using a dummy response matrix based on a binary coding. The class assignment is based on a simple dichotomous distinction such as breast cancer (class 1) / healthy (class 2), or a multiple distinction based on multiple disease diagnosis such as breast cancer (class 1) / Alzheimer (class 2) / healthy (class 3). The list of diseases for classification can be increased depending upon the samples available corresponding to other diseases or conditions or stages thereof.

PLSR applied as a classification method is referred to as PLS-DA (DA standing for Discriminant analysis).

PLS-DA is an extension of the PLSR algorithm in which the Y-matrix is a dummy matrix containing n rows (corresponding to the number of samples) and K columns (corresponding to the number of classes). The Y-matrix is constructed by inserting 1 in the k th column and -1 in all the other columns if the corresponding i th object of X belongs to class k . By regressing Y onto X , classification of a new sample is achieved by selecting the group corresponding to the largest component of the fitted, $\hat{Y}(x) = (\hat{Y}_1(x), \hat{Y}_2(x), \dots, \hat{Y}_K(x))$.

Thus, in a -1/1 response matrix, a prediction value below 0 means that the sample belongs to the class designated as -1, while a prediction value above 0 implies that the sample belongs to the class designated as 1.

An advantage of PLSR-DA is that the results obtained can be easily represented in the form of two different plots, the score and loading plots. Score plots represent a projection of the samples onto the principal components and shows the distribution of the samples in the classification model and their relationship to one another. Loading plots display correlations between the variables present in the data

set.

It is usually recommended to use PLS-DA as a starting point for the classification problem due to its ability to handle collinear data, and the property of PLSR as a dimension reduction technique. Once this purpose has been satisfied, it is possible to use other methods such as Linear discriminant analysis, LDA, that has been shown to be effective in extracting further information, Indahl et al. (1999, Chem. and Intell. Lab. Syst., 49, p19-31). This approach is based on first decomposing the data using PLS-DA, and then using the scores vectors (instead of the original variables) as input to LDA. Further details on LDA can be found in Duda and Hart (Classification and Scene Analysis, 1973, Wiley, USA).

The next step following model building is of model validation. This step is considered to be amongst the most important aspects of multivariate analysis, and tests the "goodness" of the calibration model which has been built. In this work, a cross validation approach has been used for validation. In this approach, one or a few samples are kept out in each segment while the model is built using a full cross-validation on the basis of the remaining data. The samples left out are then used for prediction/classification. Repeating the simple cross-validation process several times holding different samples out for each cross-validation leads to a so-called double cross-validation procedure. This approach has shown to work well with a limited amount of data, as is the case in some of the Examples described here. Also, since the cross validation step is repeated several times the dangers of model bias and overfitting are reduced.

Once a calibration model has been built and validated, genes exhibiting an expression pattern that is most relevant for describing the desired information in the model can be selected by techniques described in

the prior art for variable selection, as mentioned elsewhere. Variable selection will help in reducing the final model complexity, provide a parsimonious model, and thus lead to a reliable model that can be used for prediction. Moreover, use of fewer genes for the purpose of providing diagnosis will reduce the cost of the diagnostic product. In this way informative probes which would bind to the genes of relevance may be identified.

We have found that after a calibration model has been built, statistical techniques like Jackknife (Effron, 1982, The Jackknife, the Bootstrap and other resampling plans. Society for Industrial and Applied mathematics, Philadelphia, USA), based on resampling methodology, can be efficiently used to select or confirm significant variables (informative probes).

The approximate uncertainty variance of the PLS regression coefficients B can be estimated by:

$$S^2B = \sum_{m=1}^M ((B-B_m)g)^2$$

where

S^2B = estimated uncertainty variance of B;

B = the regression coefficient at the cross validated rank A using all the N objects;

B_m = the regression coefficient at the rank A using all objects except the object(s) left out in cross

validation segment m; and

g = scaling coefficient (here: g=1).

In our approach, Jackknife has been implemented together with cross-validation. For each variable the difference between the B-coefficients B_i in a cross-validated sub-model and B_{tot} for the total model is first calculated. The sum of the squares of the differences is then calculated in all sub-models to

obtain an expression of the variance of the B_1 estimate for a variable. The significance of the estimate of B_1 is calculated using the t-test. Thus, the resulting regression coefficients can be presented with
5 uncertainty limits that correspond to 2 Standard Deviations, and from that significant variables are detected.

No further details as to the implementation or use of this step are provided here since this has been
10 implemented in commercially available software, The Unscrambler (CAMO ASA, Oslo, Norway). Also, details on variable selection using Jackknife can be found in Westad & Martens (2000, J. Near Inf. Spectr., 8, p117-124).

15 The following approach can be used to select informative probes from a gene expression data set:

a) keep out one unique sample (including its repetitions if present in the data set) per cross validation segment;

20 b) build a calibration model (cross validated segment) on the remaining samples using PLSR-DA;

c) select the significant genes for the model in step b) using the Jackknife criterion;

d) repeat the above 3 steps until all the unique
25 samples in the data set are kept out once (as described in step a). For example, if 75 unique samples are present in the data set, 75 different calibration models are built resulting in a collection of 75 different sets of significant probes;

30 e) select the most significant variables using the frequency of occurrence criterion in the generated sets of significant probes in step d). For example, a set of probes appearing in all sets (100%) are more informative than probes appearing in only 50% of the generated sets
35 in step d).

Once the informative probes for a disease have been selected, a final model is made and validated. The two

most commonly used ways of validating the model are cross-validation (CV) and test set validation. In cross-validation, the data is divided into k subsets. The model is then trained k times, each time leaving out one of the subsets from training, but using only the omitted subset to compute error criterion, RMSEP (Root Mean Square Error of Prediction). If k equals the sample size, this is called "leave-one-out" cross-validation. The idea of leaving one or a few samples out per validation segment is valid only in cases where the covariance between the various experiments is zero. Thus, one sample at-a-time approach can not be justified in situations containing replicates since keeping only one of the replicates out will introduce a systematic bias in our analysis. The correct approach in this case will be to leave out all replicates of the same samples at a time since that would satisfy assumptions of zero covariance between the CV-segments.

The second approach for model validation is to use a separate test-set for validating the calibration model. This requires running a separate set of experiments to be used as a test set. This is the preferred approach given that real test data are available. This requires keeping one (or several samples including replicates) out as a test sample. The remaining samples are then used for building the calibration model using the simple cross-validation routine explained above. Hence, the name double cross-validation. Once the calibration model has been built it is then used to predict the sample(s) kept out as the test sample(s). This procedure is continued until each and every sample is kept out and used as a test sample.

The final model is then used to identify a disease, condition or stage thereof in test samples. For this purpose, expression data of selected informative genes is generated from test samples and then the final model is used to determine whether a sample belongs to a

diseased or non-diseased class or has a condition or stage thereof.

Thus viewed from a yet further aspect the present invention provides a method of identifying probes useful for diagnosing or identifying or monitoring a disease or condition or stage thereof in an organism, comprising the steps of:

- a) immobilizing a set of oligonucleotide probes, preferably as described hereinbefore, on a solid support;
- b) isolating mRNA from a sample of a normal organism (normal sample), which may optionally be reverse transcribed to cDNA;
- c) isolating mRNA from a sample from an organism, corresponding to the sample and organism of step (b), which is known to have said disease or condition or a stage thereof (diseased sample), which may optionally be reverse transcribed to cDNA;
- d) hybridizing the mRNA or cDNA of steps (b) and (c) to said set of immobilized oligonucleotide probes of step (a); and
- e) assessing the amount of mRNA or cDNA hybridizing to each of said oligonucleotide probes to determine the level of gene expression of genes to which said oligonucleotide probes bind in said normal and diseased samples to generate a gene expression data set for each sample;
- f) normalizing and standardizing said data set of step (e);
- g) constructing a calibration model for classification, preferably using the statistical techniques Partial Least Squares Discriminant Analysis (PLS-DA) and Linear Discriminant Analysis (LDA);
- h) performing JackKnife analysis and identifying

those oligonucleotide probes which are required for classification of said disease and normal samples into their respective groups.

5 Preferably a model for classification purposes is generated by using the data relating to the probes identified according to the above described method. Preferably the sample is as described previously. Preferably the oligonucleotides which are immobilized in
10 step (a) are randomly selected as described below or are the probes as described hereinbefore. Such oligonucleotides may be of considerable length, e.g. if using cDNA (which is encompassed within the scope of the term "oligonucleotide"). The identification of such
15 cDNA molecules as useful probes allows the development of shorter oligonucleotides which reflect the specificity of the cDNA molecules but are easier to manufacture and manipulate.

 The above described model may then be used to
20 generate and analyse data of test samples and thus may be used for the diagnostic methods of the invention. In such methods the data generated from the test sample provides the gene expression data set and this is normalized and standardized as described above. This is
25 then fitted to the calibration model described above to provide classification.

 The method described herein can also be used to simultaneously select informative probes for several related and unrelated diseases or conditions. Depending
30 upon which diseases or conditions have been included in the calibration or training set, informative probes can be selected for the said diseases or conditions. The informative probes selected for one disease or condition may or may not be similar to the informative probes
35 selected for another disease or condition of interest. It is the pattern with which the selected genes are expressed in relation to each other during a disease,

condition, or stage thereof, that determines whether or not they are informative for the disease, condition or stage thereof.

5 In other words, informative genes are selected
based on how their expression correlates with the
expression of other selected informative genes under the
influence of responses generated by the disease,
condition or stage thereof under investigation. In the
examples provided hereinafter, 139 informative probes
10 were selected for breast cancer diagnosis and 182 probes
were selected for Alzheimer's disease diagnosis by
training the gene expression data set of genes
representing 1435 or 758 randomly picked cDNA clones for
breast cancer/non breast cancer samples, or
15 Alzheimer/non-Alzheimer samples, respectively. Among
the probes selected for breast cancer and Alzheimer,
about 10 probes were informative both for breast cancer
and Alzheimer disease diagnosis.

For the purpose of isolating informative probes or
20 identifying several related and unrelated diseases,
conditions and stages thereof simultaneously, the gene
expression data set must contain the information on how
genes are expressed when the subject has a particular
disease, condition or stage thereof under investigation.
25 The data set is generated from a set of healthy or
diseased samples, where a particular sample may contain
the information of only one disease, condition or stages
thereof or may also contain information about multiple
diseases, conditions or stages thereof. For example, if
30 the isolation of informative probes for Alzheimer
disease, breast cancer and diabetes is sought, whole
blood samples can be obtained from an Alzheimer patient
who has breast cancer and diabetes. Hence, the method
also teaches an efficient experimental design to reduce
35 the number of samples required for isolating informative
probes by selecting samples representing more than one
disease, condition or stage thereof.

As mentioned previously, in view of the high information content of most transcripts, the identification and selection of informative probes for use in diagnosing, monitoring or identifying a particular disease, condition or stage thereof may be dramatically simplified. Thus the pool of genes from which a selection may be made to identify informative probes may be radically reduced.

Unlike, in prior art technologies where informative probes are selected from a population of thousands of genes that are being expressed in a cell, like in microarray, in the method described herein, the informative probes are selected from a limited number of randomly obtained genes. For example, from a population of 1435 cDNA clones, randomly picked from a human whole blood cDNA library, we were able to select 182 informative probes for breast cancer diagnosis (see the Examples).

Thus in a preferred aspect of the above mentioned method of identifying probes useful for diagnosing or identifying or monitoring a disease or condition or stage thereof in an organism, said set of oligonucleotides which are immobilized in step (a) are randomly selected from a larger set of oligonucleotides, e.g. from a cDNA library or other oligonucleotide pool. Preferably said larger set comprises oligonucleotides which correspond to moderately or highly expressed genes.

As referred to herein "random" refers to selection which is not biased based on the extent of information carried by the transcripts in relation to the disease, condition or organism under study, ie. without bias towards their likely utility as informative probes. Whilst a random selection may be made from a pool of transcripts (or related products) which have been biased, e.g. to highly or moderately expressed transcripts, preferably random selection is made from a

pool of transcripts not biased or selected by a sequence-based criterion.

The larger set may therefore contain oligonucleotides corresponding to highly and moderately expressed genes, or alternatively, may be enriched for those corresponding to the highly and moderately expressed genes.

Random selection from highly and moderately expressed genes can be achieved in a wide variety of ways. A strategy used in this work, but not limiting in itself involves randomly picking a significant number of cDNA clones from a cDNA library constructed from a biological specimen under investigation. Since, in a cDNA library, the cDNA clones corresponding to transcripts present in high or moderate amount are more frequently present than transcripts corresponding to cDNA present in low amount, the former will tend to be picked up more frequently than the latter. A pool of cDNA enriched for those corresponding to highly and moderately expressed genes can be isolated by this approach.

To identify genes that are expressed in high or moderate amount among the isolated population for use in methods of the invention, the information about the relative level of their transcripts in samples of interest can be generated using several prior art techniques. Both non-sequence based methods, such as differential display or RNA fingerprinting, and sequence-based methods such as microarrays or macroarrays can be used for the purpose. Alternatively, specific primer sequences for highly and moderately expressed genes can be designed and methods such as quantitative RT-PCR can be used to determine the levels of highly and moderately expressed genes. Hence, a skilled practitioner may use a variety of techniques which are known in the art for determining the relative level of mRNA in a biological sample.

Especially preferably the sample for the isolation of mRNA in the above described method is as described previously and is preferably not from the site of disease and the cells in said sample are not disease cells and have not contacted disease cells.

The following examples are given by way of illustration only in which the Figures referred to are as follows:

10

Figure 1 shows the effect of Direct Standardization (DS) on the Alzheimer data measured in two different series of experiments;

15

Figure 2 shows the projection of normal (including benign) and breast cancer samples onto a classification model generated by PLSR-DA using the data of 44 informative genes; and

20

Figure 3 shows the projection of individuals with and without Alzheimer's disease onto a classification model generated by PLSR-DA using 182 informative genes.

Example 1: Diagnosis of Breast Cancer

Methods

25

Whole blood was obtained from the arms of breast cancer patients and patients with benign tumours (Ullevål and Haukland hospitals in Norway). All of the patients with breast cancer had a malignant tumour of the breast (disease samples). Healthy blood was collected from the above two hospitals, or collected at a Health station at ... Ås, Norway or at DiaGenic AS, Norway, from the arms of female donors with no reported signs of breast cancer. The blood from healthy individuals or with benign tumours comprise the normal samples. The blood was either collected in tubes containing EDTA and stored immediately at -80°C or was collected in PAXgene tubes

35

and stored for 12-24 hours at room temperature before finally storing them at -80°C before use. Further details of the breast cancer and benign tumour patients from which blood was taken is provided in Table 5.

5

mRNA was isolated from the blood of the 29 breast cancer patients and 46 normal donors and used to prepare labelled probes by reverse transcribing in the presence of $\alpha^{33}\text{P}$ -dATP.

10

The first strand cDNA of the normal and diseased samples was bound, separately to 1435 cDNA clones immobilized on a solid support (nylon membrane). These cDNA clones were randomly picked, without any prior knowledge of their gene sequences, from a cDNA library constructed using whole blood of 550 healthy individuals. cDNAs were amplified from these clones, and the amplified products denatured and spotted on the membrane using Biorobotics. In addition to 1435 cDNAs, several types of controls such as negative controls, different external controls, and cancer markers were also spotted on to the membrane.

20

The amount of labelled first strand cDNA binding to each spot was assessed and quantified using a PhosphoImager to generate a gene expression data set. The data was generated using Phoretix software version 3 (Non Linear Dynamics, England). Background subtraction was performed on the generated data by subtracting the median of the line of pixels around each spot outline from the total intensity obtained from the respective spots.

30

The background-subtracted data was then normalized and transformed by selecting out 50 lowest and 50 maximum signals from each membrane. This step was to exclude genes that were expressed with a high degree of

35

variance. Since the genes varied from membrane to membrane, the expression data from 497 genes were removed from the data set. The values for the remaining 938 genes were then normalised by using different approaches such as external controls, dividing each spot by the median intensity of the observed signal in the respective membrane, range normalizing the data from each membrane, and then log transforming the data obtained.

The processed data obtained above was then used to isolate the informative probes by:

a) keeping one unique sample (including all repetitions of the selected sample) out per cross validation segment;

b) building a calibration model (cross validated) on the remaining samples using PLSR-DA;

c) selecting the set of significant genes for the model in step b using the Jackknife criterion;

d) repeating steps a), b) and c) until all the unique samples were kept out once (hence, in all 75 different calibration models were built (after repeating step b) 75 times), resulting in 75 different sets of significant probes (after repeating step c) 75 times));

e) selecting significant variables using the frequency of occurrence criterion amongst the 75 different sets of significant probes.

The selected informative probes based on occurrence criterion were used to construct a classification model. The result of the classification model based on probes appearing in at least 90% of the generated sets after the step of isolating informative probes as described above is shown in Figure 2 in which it is seen that the expression pattern of these genes was able to classify most women with breast cancer and women with no breast cancer into distinct groups. In this figure PC1 and PC2

indicate the two principal components statistically derived from the data which best define the systemic variability present in the data. This allows each sample, and the data from each of the informative probes to which the sample's labelled first strand cDNA was bound, to be represented on the classification model as a single point which is a projection of the sample onto the principal components - the score plot.

The ability of the generated model, based on isolated informative probes, to predict future samples was determined by the double cross-validation approach. The performance of the diagnostic test for breast cancer based on the occurrence criterion is presented in Table 6.

Example 2: Diagnosis of Alzheimer's disease

Similar experiments were conducted with samples from Alzheimer's patients. In this method 7 patients diagnosed with Alzheimer's Disease at the Memory Clinic at Ullevål University Hospital were used in the trial. The patients were confirmed as having Alzheimer's disease based on the following criteria:

- * A standardized interview with a care-giver using IQCODE, an ADL scale and a scale measuring behaviour of the patient (Green scale).
- * Neuropsychological evaluation using MMSE, Clock drawing test, Trailmaking test A and B (TMT A and B), Kendrick object learning test (visual memory test), part of the Wechsler battery and Benton test.
- * A psychiatric evaluation using scales for detection of depression, MADRS for interviewing the patient and Cornell scale for interviewing the care-giver.
- * A physical examination.
- * Laboratory tests of blood samples to rule out other

diseases.

- * CT scan of the brain.
- * SPECT of the brain.

5 The mean age of the patients was 72.3 with an age range of 69-76. The mean MMSE score was 22.0 (the maximum score attainable being 30).

10 Six age-matched individuals without diagnosed Alzheimer's disease were used as a control. All had been tested with MMSE and had a minimum score of 28 (mean: 28.4). The mean age of the normal control group was 73.0 and the age range 66-81. A sample from a 16-
15 year old individual, with a consequent minimal chance of having Alzheimer's disease, was also included as an additional control.

20 Using the methods described above (except that hybridization to 758 rather than 1435 cDNA clones was performed), informative probes were selected based on occurrence criterion and used to construct a
25 classification model. The results of the classification model based on probes appearing at least once in the generated sets after the method to isolate informative probes as described above is shown in Figure 3 in which
30 it will be seen that the expression pattern of these genes was able to classify individuals with or without Alzheimer's disease into distinct groups. In this Figure PC1 and PC2 indicate the 2 principal components statistically derived from the data which define the
35 systematic variability present in the data. This allows each sample, and the data from each of the informative probes to which the samples' cDNA was bound, to be represented on the classification model as a single point which is a projection of the sample onto the principal components - the score plot.

The ability of the generated model, based on isolated informative probes, to predict future samples was determined by the double cross-validation. The performance of the diagnostic test for Alzheimer's disease is presented in Table 7.

5

Appendix A

Partial Least Squares regression (PLSR)

5

Let a multivariate regression model be defined as:

$$Y = XB + F$$

10

where

X a $N \times P$ matrix with N predictor variables (genes);
Y ($N \times J$) being the J predicted variables. In our case Y
represents a matrix containing dummy variables;
B is a matrix of regression coefficients; and

15

F is a $N \times J$ matrix of residuals.

The structure of the PLSR model can be written as:

$$X = TP^T + E_A, \text{ and}$$

20

$$Y = TQ^T + F_A, \text{ where}$$

where

T ($N \times A$) is a matrix of score vectors which are linear
combinations of the x-variables;

25

P ($P \times A$) is a matrix with the x-loading vectors p_a as
columns;

Q ($J \times A$) is a matrix with the y-loading vectors q_a as
columns;

E_A ($N \times P$) is the matrix for X after A factors; and

30

F_A ($N \times J$) is the matrix for Y after A factors.

The criterion in PLSR is to maximize the explained
covariance of $[X, Y]$. This is achieved by the loading
weights vector w_{a+1} , which is the first eigenvector of
35 $E_A^T F_A F_A^T E_A$ (E_A and F_A are the deflated X and Y after a
factors or PLS components).

The regression coefficients are given by:

$$B = W(P^TW)^{-1}Q^T$$

- 5 A PLSR model with full rank, i.e. maximum number of components; is equivalent to the MLR solutions. Further details on PLSR can be found in Martens & Naes, 1989, Multivariate Calibration, John Wiley & Sons, Inc., USA; and Kowalski & Seasholtz, 1991, supra.

- 65 -

Table 1

List of probes informative for disease diagnosis

	Clone ID	Sequence ID	No. of nucleotides
1	I-01	-	-
2	I-02	-	-
3	I-13	-	-
4	I-21	-	-
5	I-24	308	373
6	I-28	310	564
7	I-30	1180	622
8	I-34	313	554
9	I-37	-	-
10	I-42	-	-
11	I-52	-	-
12	I-54	1181	155
13	I-58	326	554
14	I-71	-	-
15	I-72	-	-
16	I-86	-	-
17	I-95	-	-
18	II-03	361	622
19	II-05	363	628
20	II-06	364	528
21	II-10	368	329
22	II-24	381	534
23	II-25	382	444
24	II-26	383	566
25	II-33	390	523
26	II-34	391	566
27	II-41	397	534
28	II-42	398	512
29	II-47	-	-
30	II-57	411	505
31	II-61	415	596
32	II-69	423	387
33	II-70	424	420
34	II-75	429	535
35	II-83	-	-
36	II-84	438	577
37	II-87	441	552
38	II-88	442	606
39	II-90	-	-
40	II-94	448	329
41	III-02	453	747
42	III-05	-	-
43	III-06	458	682

44	III-08	460	536
45	III-10	-	-
46	III-13	464	615
47	III-15	-	-
48	III-17	-	-
49	III-20	1183	479
50	III-23	473	694
51	III-26	476	476
52	III-35	485	551
53	III-39	487	224
54	III-40	488	349
55	III-43	490	382
56	III-44	491	382
57	III-53	500	390
58	III-56	503	109
59	III-57	504	374
60	III-60	-	-
61	III-60	-	-
62	III-61	507	521
63	III-63	509	575
64	III-68	-	-
65	III-74	518	502
66	III-80	523	585
67	III-82	-	-
68	III-85	526	516
69	III-89	530	660
70	III-92	-	-
71	III-96	-	-
72	IV-14	684	545
73	IV-15	1185	628
74	IV-23	-	-
76	IV-26	1186	494
75	IV-26	-	-
77	IV-29	-	-
78	IV-31	687	268
79	IV-32	688	569
80	IV-34	-	-
81	IV-35	-	-
82	IV-41	-	-
83	IV-45	-	-
84	IV-53	61	362
85	IV-62	-	-
86	IV-69	192	286
87	IV-80	701	579
88	IV-82	-	-
89	IV-93	-	-
90	IX-10	736	641
91	IX-12	-	-
92	IX-38	757	583
93	IX-39	758	424
94	IX-42	-	-
95	IX-48	764	626
96	IX-77	785	556
97	V-01	-	-
98	V-02	-	-
99	V-03	706	496

100	V-04	707	397
101	V-06	-	-
102	V-07	708	293
103	V-11	1188	599
104	V-12	711	498
105	V-15	-	-
106	V-17	-	-
107	V-21	-	-
108	V-25	-	-
109	V-32	-	-
110	V-35	-	-
111	V-39	-	-
112	V-42	-	-
113	V-43	-	-
114	V-47	-	-
115	V-49	-	-
116	V-52	-	-
117	V-54	-	-
118	V-55	77	412
119	V-58	-	-
120	V-59	-	-
121	V-65	-	-
122	V-68	-	-
123	V-71	-	-
124	V-75	-	-
125	V-79	-	-
126	V-80	726	260
127	V-90	-	-
128	V-91	-	-
129	V-92	-	-
130	V-94	-	-
131	VI-02	-	-
132	VI-04	865	122
133	VI-07	93	405
134	VI-09	-	-
135	VI-10	-	-
136	VI-12	869	667
137	VI-14	871	642
138	VI-17	-	-
139	VI-20	876	115
140	VI-21	-	-
141	VI-23	878	634
142	VI-34	-	-
143	VI-41	-	-
144	VI-42	-	-
145	VI-43	-	-
146	VI-44	-	-
147	VI-48	891	626
148	VI-49	-	-
149	VI-50	893	585
150	VI-52	-	-
151	VI-53	895	560
152	VI-55	897	509
153	VI-65	-	-
154	VI-70	108	550
155	VI-71	-	-

156	VI-72	-	-
157	VI-74	905	655
158	VI-76	907	582
159	VI-78	-	-
160	VI-79	-	-
161	VI-84	-	-
162	VI-87	911	595
163	VI-88	912	651
164	VI-90	-	-
165	VI-93	-	-
166	VI-95	915	230
167	VI-96	-	-
168	VII-02	-	-
169	VII-03	1196	412
170	VII-06	-	-
171	VII-10	-	-
172	VII-11	-	-
173	VII-15	1199	439
174	VII-19	562	580
175	VII-21	564	671
176	VII-25	-	-
177	VII-32	571	457
178	VII-36	575	209
179	VII-39	576	541
180	VII-42	579	502
181	VII-43	580	316
182	VII-46	583	631
183	VII-47	1200	526
184	VII-48	1201	613
185	VII-59	593	565
186	VII-60	-	-
187	VII-63	595	98
188	VII-66	598	362
189	VII-67	-	-
190	VII-72	600	595
191	VII-73	601	522
192	VII-75	-	-
193	VII-76	603	624
194	VII-77	1203	692
195	VII-80	605	338
196	VII-81	606	556
197	VII-83	-	-
198	VII-86	-	-
199	VII-88	-	-
200	VII-90	612	576
201	VII-91	613	341
202	VII-93	615	379
203	VIII-01	-	-
204	VIII-02	-	-
205	VIII-03	-	-
206	VIII-06	-	-
207	VIII-09	618	598
208	VIII-10	-	-
209	VIII-15	-	-
210	VIII-20	628	419
211	VIII-22	-	-

212	VIII-26	-	-
213	VIII-28	634	511
214	VIII-29	635	592
215	VIII-30	636	572
216	VIII-31	637	482
217	VIII-32	638	545
218	VIII-33	639	624
219	VIII-39	-	-
220	VIII-41	645	649
221	VIII-42	646	600
222	VIII-44	-	-
223	VIII-46	649	425
224	VIII-48	651	251
225	VIII-58	-	-
226	VIII-64	663	627
227	VIII-65	-	-
228	VIII-66	665	345
229	VIII-67	666	252
230	VIII-74	-	-
231	VIII-76	675	591
232	VIII-78	-	-
233	VIII-82	-	-
234	VIII-83	-	-
235	VIII-85	-	-
236	VIII-87	-	-
237	VIII-91	-	-
238	VIII-92	-	-
239	VIII-93	-	-
240	VIII-95	-	-
241	X-04	-	-
242	X-07	808	641
243	X-15	814	132
244	X-29	821	370
245	X-34	-	-
246	X-35	-	-
247	X-54	837	603
248	X-56	839	71
249	X-68	1207	642
250	X-72	849	622
251	X-94	860	501
252	XI-07	-	-
253	XI-13	1209	620
254	XI-50	-	-
255	XI-58	-	-
256	XI-81	1212	374
257	XII-07	1213	567
258	XII-17	-	-
259	XII-26	-	-
260	XII-27	-	-
261	XII-31	-	-
262	XII-32	-	-
263	XII-35	1214	620
264	XII-36	-	-
265	XII-52	-	-
266	XII-59	1216	484
267	XIII-19	1219	559

268	XIII-29	-	-
269	XIII-52	939	513
270	XIII-62	-	-
271	XIII-84	-	-
272	XIII-92	1221	741
273	XV-18	-	-
274	XV-22	1099	561
275	XV-24	-	-
276	XV-25	1224	485
277	XV-28	-	-
278	XV-34	-	-
279	XV-42	-	-
280	XV-68	-	-
281	XV-74	-	-
282	XV-93	-	-
283	XV-94	-	-
284	XV-96	-	-
285	XVI-36	1056	435
286	XVI-53	1230	741
287	XVI-59	-	-
288	XVI-66	1074	689
289	XVI-76	1083	198
290	XVI-77	1084	198
291	XVII-07	-	-
292	XVII-08	-	-
293	XVII-17	-	-
294	XVII-28	-	-
295	XVII-29	-	-
296	XVII-31	1139	503
297	XVII-36	-	-
298	XVII-39	-	-
299	XVII-40	1231	203
300	XVII-48	1148	587
301	XVII-55	-	-
302	XVII-58	-	-
303	XVII-67	-	-
304	XVII-72	-	-
305	XVII-76	1160	650
306	XVII-82	-	-
307	XVII-87	1165	502
308	XVII-95	1172	648

-69-

Table 2

List of informative probes for diagnosis of breast cancer

Clone ID	Sequence ID
I-24	308
I-28	310
I-30	1180
I-52	-
I-54	1181
II-41	397
II-70	424
II-87	441
III-06	458
III-20	1183
III-40	488
III-57	504
III-60	-
III-61	507
III-89	530
IV-14	684
IV-15	1185
IV-26	1186
IV-32	688
IV-41	-
IV-53	61
IV-62	-
IV-69	192
IV-80	701
IV-82	196
IX-10	736
IX-12	-
IX-38	757
IX-39	758
IX-42	-
IX-48	764
IX-77	785
V-11	1188
V-32	-
V-39	-
V-55	77
V-80	726
V-94	-
VI-07	93
VI-34	-
VI-41	-
VI-48	891
VI-49	-
VI-52	-
VI-55	897
VI-65	-
VI-70	108

Clone ID	Sequence ID
VI-72	-
VI-78	-
VI-84	-
VII-03	1196
VII-15	1199
VII-32	571
VII-39	576
VII-47	1200
VII-48	1201
VII-60	-
VII-73	601
VII-77	1203
VII-90	612
VIII-20	628
VIII-29	635
VIII-30	636
VIII-31	637
VIII-39	-
VIII-44	-
VIII-46	649
VIII-48	651
VIII-66	665
VIII-74	-
VIII-76	675
X-04	-
X-07	808
X-15	814
X-29	821
X-34	-
X-35	-
X-54	837
X-56	839
X-68	1207
X-72	849
X-94	860
XI-07	-
XI-13	1209
XI-50	-
XI-58	-
XI-81	1212
XII-07	1213
XII-17	-
XII-26	-
XII-27	-
XII-31	-
XII-32	-
XII-35	1214

Clone ID	Sequence ID
XII-36	-
XII-52	-
XII-59	1216
XIII-19	1219
XIII-29	-
XIII-52	939
XIII-62	-
XIII-84	-
XIII-92	1221
XV-18	-
XV-22	1099
XV-24	-
XV-25	1224
XV-28	-
XV-34	-
XV-42	-
XV-68	-
XV-74	-
XV-93	-
XV-94	-
XV-96	-
XVI-36	1056
XVI-53	1230
XVI-59	-
XVI-66	1074
XVI-76	1083
XVI-77	1084
XVII-07	-
XVII-08	-
XVII-17	-
XVII-28	-
XVII-29	-
XVII-31	1139
XVII-36	-
XVII-39	-
XVII-40	1231
XVII-48	1148
XVII-55	-
XVII-58	-
XVII-67	-
XVII-72	-
XVII-76	1160
XVII-82	-
XVII-87	1165
XVII-95	1172

Table 3

List of informative probes (Clone ID) selected for breast cancer diagnosis based on their occurrence criterion during variable selection

Occurrence*	Clone ID
100%	XI-8, XVI-66, VIII-66, XVI-59, VII-03, XIII-19, XII-35, X-35, XI-50, XII-26, IV-53, XIII-29, XIII-62, I-30, III-06, XV-22, XV-94, VII-15, VII-39, IX-39, XVII-39, III-40, VII-32
90%	I-52, VI-65, VI-34, IV-62, XV-34, XVII-58, V-11, VI-78, XII-36, XIII-92, VIII-29, XVI-53, XVI-77, XI-13, XIII-84, IV-14, XII-31, V-80, VII-48, XVII-29, XVII-72
80%	III-60, VIII-74, IX-12, X-04, XIII-52, VIII-30, IX-38
70%	VI-49, X-29, VIII-48
60%	IV-82, IX-10, VI-52, X-68, VII-77
50%	IV-15
40%	XV-28, II-70, V-55
30%	XVII-17, XVII-67
20%	XI-58, XVI-36, VIII-39, VIII-44, III-61, IV-69, XV-68, X-72
10%	IX-42, IX-77, X-94, XV-96, XVII-55
5%	XII-59, XVI-76, I-54, XV-18, V-94, X-54, VI-07, VII-47, XVII-31, XVII-87, XVII-48
In at least one model	II-41, VI-41, III-57, III-89, VII-73, XV-25, IV-26, X-34, IV-41, VII-90, XV-42, XVII-82, XII-27, VIII-20, I-28, VII-60, VIII-76, III-20, VI-84, XI-07, XVII-28, XII-17, XVII-36, XII-52, XVII-76, VIII-46, VI-70, XV-74, XV-93, VIII-31, II-87, V-39, VI-55, X-07, X-15, XII-07, XVII-07, XVII-08, XVII-95, I-24, IV-32, V-32, VI-48, VI-72, IV-80, IX-48, X-56, XV-24, XII-32, XVII-40

*100% = Genes appearing in all the 75 cross validated models; 90% = Additional genes appearing in at least 68 out of 75 cross validated models; 5% = Additional genes appearing in at least 4 out of 75 cross validated models and so on.

Table 4

List of informative probes for diagnosis of Alzheimer disease

Clone ID	Sequence ID
I-01	-
I-02	-
I-13	-
I-21	-
I-34	313
I-37	-
I-42	-
I-58	326
I-71	-
I-72	-
I-86	-
I-95	-
II-03	361
II-05	363
II-06	364
II-10	368
II-24	381
II-25	382
II-26	383
II-33	390
II-34	391
II-42	398
II-47	-
II-57	411
II-61	415
II-69	423
II-75	429
II-83	-
II-84	438
II-88	442
II-90	-
II-94	448
III-02	453
III-05	-
III-06	458
III-08	460
III-10	-
III-13	464
III-15	-
III-17	-
III-23	473
III-26	476
III-35	485
III-39	487
III-43	490
III-44	491
III-53	500
III-56	603

Clone ID	Sequence ID
III-60	-
III-63	509
III-68	-
III-74	518
III-80	523
III-82	-
III-85	526
III-92	-
III-96	-
IV-23	-
IV-26	-
IV-29	-
IV-31	687
IV-34	-
IV-35	-
IV-45	-
IV-80	701
IV-82	-
IV-93	-
V-01	-
V-02	-
V-03	706
V-04	707
V-06	-
V-07	708
V-12	711
V-15	-
V-17	-
V-21	-
V-25	-
V-35	-
V-42	-
V-43	-
V-47	-
V-49	-
V-52	-
V-54	-
V-58	-
V-59	-
V-65	-
V-68	-
V-71	-
V-75	-
V-79	-
V-80	726
V-90	-
V-91	-
V-92	-

Clone ID	Sequence ID
VI-02	-
VI-04	865
VI-09	-
VI-10	-
VI-12	869
VI-14	871
VI-17	-
VI-20	876
VI-21	-
VI-23	878
VI-41	-
VI-42	-
VI-43	-
VI-44	-
VI-48	891
VI-49	-
VI-50	893
VI-53	895
VI-71	-
VI-74	905
VI-76	907
VI-78	-
VI-79	-
VI-87	911
VI-88	912
VI-90	-
VI-93	-
VI-95	915
VI-96	-
VII-02	-
VII-03	-
VII-06	-
VII-10	-
VII-11	-
VII-19	562
VII-21	564
VII-25	-
VII-36	575
VII-42	579
VII-43	580
VII-46	583
VII-59	593
VII-63	595
VII-66	598
VII-67	-
VII-72	600
VII-73	601
VII-75	-
VI-02	-
VI-04	866
VI-09	-
VI-10	-
VI-12	873
VI-14	875
VI-17	-

Clone ID	Sequence ID
VII-91	613
VII-93	615
VIII-01	-
VIII-02	-
VIII-03	-
VIII-06	-
VIII-09	618
VIII-10	-
VIII-15	-
VIII-22	-
VIII-26	-
VIII-28	634
VIII-30	636
VIII-32	638
VIII-33	639
VIII-41	645
VIII-42	646
VIII-48	651
VIII-58	-
VIII-64	663
VIII-65	-
VIII-67	666
VIII-78	-
VIII-82	-
VIII-83	-
VIII-85	-
VIII-87	-
VIII-91	-
VIII-92	-
VIII-93	-
VIII-95	-

Table 5

Details of the breast cancer and benign tumour patients

Samples

Diagnosis	No. of women
Normal/benign	42*
DCIS	3
Invasive cancer	26

* From one woman, whole blood was collected at weeks 1,2,3,4,5 following menstruation. Hence, the number of unique normal/benign samples tested in the experiment is 75.

Tumour size in women with breast cancer

Tumour size (mm)	No. of women
4-10	11
11-20	5
21-30	3
31-40	1
41-50	1
51-60	0
61-70	1
not known	7

Age distribution

Age	No. of women
<40	8
40-49	21
50-59	18
60-69	15
>69	1
not known	8

Other diseases /conditions present in the women tested

Disease/condition	No. of women with the disease
Diabetes	3
Asthma	2
Ulcerous colitis	1
Hemochromatose	1
Crohn's disease	1
Fibromyalgia	1
Psoriasis	1
Atopic eczema	1
Rheumatism	3
Allergies	26

Prior history of cancer in the women tested

Cancer type	No. of women
Breast	3
Colon	2
Stomach	1
Skin	1

Table 6

Number of samples tested by double cross validation and success of the diagnostic test for breast cancer based on selected informative genes

Number of samples tested by double cross validation

Number of unique samples tested	75
Number of unique non cancer samples tested	48
Number of cancer samples tested	29

Success of the diagnostic test for breast cancer based on selected informative genes

Occurrence in percentage [*]	Number of Informative probes	Specificity	Sensitivity	Accuracy	False Positive rate	False negative rate	Total error rate
100.00	23	84.78	75.86	81.33	15.22	24.14	18.67
90.00	44	91.30	79.31	86.67	8.70	20.69	13.33
80.00	51	86.96	79.31	84.00	13.04	20.69	16.00
70.00	54	89.13	75.86	84.00	10.87	24.14	16.00
60.00	58	89.13	75.86	84.00	10.87	24.14	16.00
50.00	59	89.13	75.86	84.00	10.87	24.14	16.00
40.00	63	89.13	75.86	84.00	10.87	24.14	16.00
30.00	66	86.96	75.86	82.67	13.04	24.14	17.33
20.00	74	89.13	75.86	84.00	10.87	24.14	16.00
10.00	79	89.13	75.86	84.00	10.87	24.14	16.00
5.00	90	86.96	79.31	84.00	13.04	20.69	16.00
1.33	139	84.78	72.41	80.00	15.22	27.59	20.00

^{*}100% = Genes appearing in all the 75 cross validated models; 90% = Genes appearing in at least 68 out of 75 cross validated models; 5% = Genes appearing in at least 4 out of 75 cross validated models; and so on.

Table 7

Double cross-validation and details of the success of the diagnostic test for Alzheimer disease based on the expression 182 informative genes

Validation Result

Total number of samples tested	14
Number of Alzheimer's disease samples tested	7
Number of Alzheimer's disease samples incorrectly predicted	1
Number of non-Alzheimer's disease samples tested	7
Number of non-Alzheimer's disease samples incorrectly predicted	0

Success of diagnostic test

Performance	Description	%
Accuracy	Percentage of the total number of predictions that were correct	82.9
Sensitivity	Percentage of positive cases that were correctly identified	85.7
Specificity	Percentage of negatives cases that were correctly predicted	100
False positive rate	Percentage of positives cases that were incorrectly classified as positive	0.0
False negative rate	Percentage of negatives cases that were incorrectly classified as negative	14.3
Total error rate	Percentage of the total cases incorrectly predicted	7.1

-77-

List of nucleotide sequences

Sequence ID - 93 nt: 405

GGATCCTGTGGCCACAGAGCTGCCCCAGCAGACGCTCCGCCCCACCCGGTGATG
GAGCCCCGGGGGACAATCGTGCCTGGGGAGGAGCAGGGTACAGCCCATTC
CAGCCCTGGCTGACCTGGCCTAGCAGTTTGGCCCTGCTGGCCTTAGCAGGGAGAC
AGGGGAGCAAAGAACGCCAAGCCGGAGGCCCCGAGGCCAGCCGGCCTCTCGAGA
GCCAGAGCAGCAGTTGAATGTAATGCTGGGGACAGGCATGCTGCCGCCAGTAGG
GCGGGGACCCGGACAGCCAGGTGACTACCAAGTCCTGGGGACACACTCACCATAA
ACACATCCCCAGGCAGGACAGATCGGGGAAGGGGTGTGTACCAGGCTATGATTT
CTCTTGCAATTAATGTATTATTATT

Sequence ID - 108 nt: 550

GGCTTTGACAGAGTGCAAGACGATGACTTGCAAAATGTCGCATCTGGAACGCAA
CATAGANACCATCATCAACACCTTCCACCAATACTCTGTGAAGCTGGGGCACCCA
GACACCCTGAACCAGGGGGAATTCAAAGAGCTGGTGCGAAAAGATCTGCAAAAT
TTTCTCAAGAAGGAGAATAAGAATGAAAAGGTCATAGAACACATCATGGAGGAC
CTGGACACAAATGCAGACAAGCAGCTGAGCTTCGAGGAGTTTATCATGCTGATG
GCGAGGCTAACCTGGGCCTCCACGAGAAGATGCACGAGGGTGACGAGGGGCCCT
GGCCACCACCATAAGCCAGGCCTCGGGGAGGGGCACCCCTAAGACCACAGTGGC
CAAGATCACAGTGGCCACGGCCACGGCCACAGTCATGGTGGCCACGGCCACAGC
CACTAATCAGGAGGCCAGGCCACCCTGCCTNTACCCAACCAGGGCCCCGGGGCCT
GTTATGTCAAACCTGTCTTGGCTGTGGGGCTAGGGGCTGGGGCCAAATAAAGTCTC
TTTCTCC

Sequence ID - 192 nt: 286

CCGGTAATAGAAATAGAAAAGGGAGAGTGTCTTCATGCAATGTGGCATCCTGGATT
GGGTCTCGNNACAAAAACAGGACATTAGTGGGAAAATTGGAAATCTGAAAAAAG
TCTGAATTTTAGTTAATATACCAATTTTCAAGTCTCTTGGTTTTGACAGATGTACCAT
GGTGATGTAAGATGTTGACCTTGGGGTAGGCTGGGTGAAGGGTATACAGGAAT
CTTTGTACTATCTCTGCAACTTCTCTGTAAATCTAGTATCATTCAAAATAAAAGT
TTATTTAATTT

Sequence ID - 308 nt: 373

AAGTGGGTCTTGCCATCCCTGAACTGNAATCATCCCTAACATATTCATACCTGTTT
TCATTTTAAAAGTTGGGTCAAGTTTTTTTATTAGTACATGTATTTCTATCCTACTGAT
TTATTTGCTATATCATCTAATTTAGTTTGAATATTCCATAATTTACTTAATTAGTCC
TGTATGGAGACCTAGCTCTTCTCAGTGTCTACTATTATAAACAATGCTACAGTGA
ATATTGGTGNATAAATCCATACNCACCACGTACATATCTTAAGTTCTGGAAGAGA
TATTGCTAAACCAGAAGATAACCTGCATTTAAAATTTGACTGCTAGGGNCAGGGN
CACATTTAATTAAATTAGAACAANGAATGCATAATGNC

Sequence ID - 310 nt:564

CCTGGNCAGAGGCCTCTATCCTGTANTGATAATTGCCATCAAAATTGTCAAAAAN
GATTTAATTTCTATGGGNAATAGTCCTTTTCTTAGCTTCTGCCNNTCACTTGCTTAT
TTTTGTGTGGGAATGGGGTTGGATAAACCAATGAACCTTATTATAAACAATCC

CACCTATATCTANCAAATTTATATTTTCGGTGAAATACAGATATTTGCCTTTCTGG
AGTANTATAGAAGCTGTCAATATGTATCTACTGTACAGTACTAAATAGTATTCAT
TTATGAAATGAGTAGTGTGTTGGGTGGCTGGGGTTAAGGAAAAATGAGACTTGGA
ATTGTAGCTTTTATCCAAGTTTTGAGTATAAATAGGGTTTTGTTTTGTTTTTTTAA
CCTAAAACTGAAATGCCATATAGAAAAACAGCATTGTTTTACAGTTTGTAGTA
AGTAACTTTTTAAAGATTTTATCAAAAAGAATTTTGTCTATNGTGAGTAAAAGAA
GTTCTAATAATGGCCTAATCACTGCATTTTTAAAAAACAAAGTTCAACACAAATG
ACATTTGTTT

Sequence ID - 397

nt: 534

GACCCGGAATCGCGGCCGCGTCGACGGAAGCTCCTGCCCCCTCTAAAGCTGAAGC
CAAAGCGAAGGCTTTAAAGGCCAAGAAGGCAGTGTTGAAAGGTGTCCACAGCCA
CAAAAAGAAGGAGATCCGCACGTACCCACCTTCCGGCGGCCGAAGACACTGCG
ACTCCGGAGACAGCCCAAATATCCTCGGAAGAGCGCTCCAGGAGAAACAAGCT
TGACCACTATGCTATCATCAAGTTTCCGCTGACCACTGAGTCTGCCATGAAGAAG
ATAGAAGACAACAACACACTTGTGTTTATTGTGGATGTTAAAGCCAACAAGCACC
AGATTAAACAGGCTGTGAAGAAGCTGTATGACATTGATGTGGCCAAGGTCAACA
CCCTGATTCCGGCCTGATGGAGAGAAGAAGGCATATGTTGACTGGCTCCTGATTA
CGATGCTTTGGATGTTGCCAACAATAATGGGATCATTTAAACTGAGTCCAGCTGC
CTAATTCTGAATATATATATATATATATATCTTTTCACCATAA

Sequence ID - 326

nt: 554

CCCGGAATCGCGGCCCGCGTCGACAACAAACCTGCATGTTCTGCACATGTATCCA
GGAACCTAAAAAAGATAGTTTGTGTGTCTTAATTGAATAATAGTAGAT
TTATAGATTAAAGATCTATGGGTTTTTAATATGGATTANAAATCTGTGGGTTTTG
ATATGGATTANAAATCTGTGGGTTTTTAATATGGATTGGAAATCTGTGGGTTTTA
ATATGGATTAAAAACATCTGTGGGTTTTTAATATGGATTAAACATCTGTGGGTTT
TTAATATGGATTAAACATCTGGGTTTTTAATATGGATTAAACATCTGTGGGTTTTT
AATATGGGTTAAAAATCAAAAGAAAATGAACTATTGCTCCAGTGCAGGAAAAT
ACAGGCAATACTGGATACAATTAGATGGTCAGGAGCGATAACCCGGTTGCCATTG
TTTGAAGAAGAGAATAAGGNGCTAGCATTCCTATCCGTAGATAATTTGACAGCTA
GGAAATAGGGGGAGTCTTCTATGTAGTTAGTGAAGGCTAAATGAACTATTATATG
C

Sequence ID - 383

nt: 566

CTTTGAAGAACTTTGCCAAATACTTTCTTACCAATCTCATGAGGAGAGGGAACAT
GCTGAGAAACTGATGAAGCTGCAGAACCAACGAGGTGGCCGAATCTTCCTTCAG
GATATCAAGAAACCAGACTGTGATGACTGGGAGAGCGGGCTGAATGCAATGGAG
TGTGCATTACATTTGGAAAAAATGTGAATCAGTCACTACTGGAAGTGCACAAAC
TGGCCACTGACAAAAATGACCCCATTTGTGTGACTTCATTGAGACACATTACCT
GAATGAGCAGGTGAAAGCCATCAAGAATTGGGTGACCACGTGACCAACTTGCG
CAAGATGGGAGCGCCCGAATCTGGCTTGGCGGAATATCTCTTTGACAAGCACACC
CTGGGAGACAGTGATAATGAAAGCTAAGCCTCGGGCTAATTTCCCATAGCCGTG
GGGTGACTTCCCTGGTCACCAAGGCAGTGCATGCATGTTGGGGTTTCCTTTACCTT
TTCTATAAGTTGTACCAAAACATCCACTTAAGTTCTTTGATTGTCCATTTCCTTCA
AATAAGAAATTTGGTA

Sequence ID - 424

nt: 420

CGCAGAATGGCTCCCGCAAAGAAGGGTGGCGAGAAGAAAAAGGGCCGTTCTGCC
ATCAACGAAGTGGTAACCCGAGAATACACCATCAACATTCAAGCGCATCCAT
GGAGTGGGCTTCAAGAAGCGTGCACCTCGGGCACTCAAAGAGATTGCGAAATTT
GCCATGAAGGAGATGGGAACCTCAGATGTGCGCATTGACACCAGGCTCAACAAA
GCTGTCTGGGCCAAAGGAATAAGGAATGTGCCATACCGAATCCGTGTGCGGCTGT
CCAGAAAACGTAATGAGGATGAAGATTCACCAAATAAGCTATATACTTTGGTTAC
CTATGTACCTGTTACCACTTTCAAAAATCTACAGACAGTCAATGTGGATGAGAAC
TAATCGCTGATCGTCAGATCAAATAAAGTTATAAAATTG

Sequence ID - 441

nt: 552

AAACAAAATTATTCTCTGAGAGGGAAAGGACATTTGAGGGAAACATCAAATTTTC
CCCATAAATAAATGAATGGAGTTTGCAGGAAGGTGAGGGTGAGCAGAGATGTGT
GTGGACATCTCTGACCATCCATCGCTGTATTCAAATGGATTGTTTTATTCCATTCT
GGTCTCAGGCATGACCACGTCCAGTGAAGACATTTGAGGCAGCACATCTCAGGAC
CCAGGCAATAGACTGGCCCCAACTCAGGCTGGACTAAGGTGTGATTAATTCTTTG
TTTTTTGTGTGGAACAGCTCACCTTGTGAGACAGCCTCAGGGCATCTCTGAGACA
CAGGGGCAGAAAATGACATTCATCTTTTGAGTCCTCATCCATGGAGTGCTGTGT
TGGGGGGCTGCATCTGCTGAAGCGAGAACCCCATCTGCCACCCACCAGGATGC
CCATTCTCAGGACTTCTCCAACCTACTATTAGACTAAACCAGAACAAAGCAACAA
ACTGTATTTATGCAAGCAAAATTGATGAGAAAATTATATTCAAATAAAGCAAAAA
TTA

Sequence ID - 458

nt: 682

TGCCACTGAAGATCCTGGTGTGCGCCATGGGGCCGCCGCCCGCCCGTTGTTACCGG
TATTGTAAGAACAAGCCGTACCCAAAGTCTCGCTTCTGCCGAGGTGTCCCTGATG
CCAAGATTTCGATTTTTTGACCTGGGGCGGAAAAAGGCAAAAGTGGATGAGTTTCC
GCTTTGTGGCCACATGGTGTGAGATGAATATGAGCAGCTGTCCTCTGAAGCCCTG
GAGGCTGCCCGAATTTGTGCCAATAAGTACATGGTAAAAAGTTGTGGCAAAGAT
GGCTTCCATATCCGGGTGCGGCTCCACCCCTTCCACGTATCCGCATCAACAAGA
TGTTGTCTGTGCTGGGGCTGACAGGCTCCAAACAGGCATGCGAGGTGCCTTTGG
AAAGCCCCAGGGCACTGTGGCCAGGGTTACATTGGCCAAGTTATCATGTCCATC
CGCACCAAGCTGCAGAACAAGGAGCATGTGATTGAGGCCCTGCGCAGGGCCAAG
TTCAAGTTTCTGGCCGCAGAAGATCCACATCTCAAAGAAGTGGGGCTTACCAAG
TTCAATGCTGATGAATTTGAAGACATGGTGGCTGAAAAGCGGCTCATCCCANATG
GCTGTGGGGTCAAGTACATCCCCAATCGTGGCCCTCTGGACAAGTGGCGGCCCTG
CACTCATGAAGGCTTTCAATGTGC

Sequence ID - 488

nt: 349

GTGCCTCCCTGTGTGAGTAGCCTAAGGTGCATTGAAAAAGACTGGGATGTGTTTT
ATTTTTTTGTATTAGATAGCATTAACTTACTGTTGAAGTATTTTTGGTGGAGTAT
TAGTGACAAGCCATTGAGTCTTAAGCCTTACGGCTTCCTATAAAATCACTAATTTC
GTGTGTGTTTGTGTGTAGGTTACGTTATATATAGGATTCGTGTTCCGCGTGGTGGC
CGAAAACGCCCAGTTCCTAAGGGTGCAACTTACGGCAAGCCTGTCCATCATGGTG
TTAACCAGCTAAAGTTTGCTCGAAGCCTTCAGTCCGTTGCAGAGGANCGAGCTGG
ACNCCCTGGGGGGCTC

Sequence ID - 504

nt: 374

CCAGCAACGACCCATACCTCAGACCCGACGGCCCGGAGCGGAGCGCGCCCTGCC
CTGGCGCAGCCAGAGCCGCGGGTGCCCGCTGCAGTTTCTTGGGACATAGGAGCG
CAAAGAAGCTACAGCCTGGACTTACCACCACTAAACTGCGAGAGAAGCTAAACG
TGTTTATTTTCCCTTAAATTATTTTGTAAATGGTAGCTTTTTCTACATCTTACTCCT
GTTGATGCAGCTAAGGTACATTTGTAAAAAGAAAAAAACCAGACTTTTCANAC
AAACCCTTTGTATTGTANATAAGAGGAAAAGACTGAGCATGCTCACTTTTTTATA
TTAATTTTACAGTATTTGTAAGAATAAAGCANCATTGAAATCG

Sequence ID - 507

nt: 521

CTGCGGTGGAGCCGCCACCAAAATGCAGATTTTCGTGAAAACCCTTACGGGGAA
GACCATCACCTCGAGGTTGAACCCTCGGATACGATAGAAAATGTAAAGGCCAA
GATCCAGGATAAGGAAGGAATTCCTCCTGATCAGCAGAGACTGATCTTTGCTGGC
AAGCAGCTGGAAGATGGACGTACTTTGTCTGACTACAATATTCAAAGGAGTCTA
CTCTTCATCTTGTGTTGAGACTTCGTGGTGGTGCTAAGAAAAGGAAGAAGAAGTC
TTACACCACTCCCAAGAAGAATAAGCACAAGAGAAAAGGAAGGTTAAGCTGGCTGT
CCTGAAATATTATAAGGTGGATGAGAATGGCAAATTAGTCGCCTTCGTGAGAG
TGCCCTTCTGATGAATGTGGTGGTGGGGTGTATGGCAAGTCACTTTGACAGAC
ATTATTGTGGCAAATGTTGTCTGACTTACTGTTTCAACAAACCAGAAGACAAGTA
ACTGTATGAGTTAATAAAAGACATGAACT

Sequence ID - 530

nt: 660

GACAGCAGAGCACACAAGCTTNTAGGACAAGAGCCAGGAAGAAACCACCGGAA
GGAACCATCTCACTGTGTGTAAACATGACTTCCAAGCTGGCCGTGGCTCTCTTGG
CAGCCTTCCTGATTTCTGCAGCTCTGTGTGAAGGTGCAGTTTTGCCAAGGAGTGCT
AAAGAAGCTTAGATGTCAGTGCATAAAGACATACTCCAACCTTTCCACCCCAAAT
TTATCAAAGAAGCTGAGAGTGATTGAGAGTGGACCACACTGCGCCAACACAGAAA
TTATTGTAAAGCTTTCTGATGGAAGANAGCTCTGTCTGGACCCCAAGGAAAAGT
GGTGCANAGGGTTGTGGANAAGTTTTTGAAGAGGGCTGAGAATTCATAAAAAAA
TTCATTCTCTGTGGTATCCAAGAATCAGTGAAGATGCCAGTGAAACTTCAAGCAA
ATCTACTTCAACACTTCATGTATTGTGTGGGTCTGTTGTAGGGTTGCCAGATGCAA
TACAAGATTCTGTTAAATTTGAATTTTCAGTAAACAATGAATAGTTTTTTCATTGT
ACCATGAAATATCCAGAACATACTTATATGTAAAGTATTATTTATTTGAATCTACA
AAAAACAACAAATAATTTTTAGATATAAGGATTTTCTGGATATTGCACGGGAGA

Sequence ID - 571

nt: 457

TTAGAGAGGTGAGGATCTGGTATTTCTGGACTAAATTCCCCTTGGGGAAGACGA
AGGGATGCTGCAGTTCCAAAAGAGAAGGACTCTTCCAGAGTCATCTACCTGAGTC
CCAAAGCTCCCTGTCTGAAAGCCACAGACAATATGGTCCCAAATGACTGACTGC
ACCTTCTGTGCCTCAGCCGTTCTTGACATCAAGAATCTTCTGTTCCACATCCACAC
AGCCAATACAATTAGTCAAACCACTGTTATTAACAGATGTAGCAACATGAGAAAC
GCTTATGTTACAGGTTACATGAGAGCAATCATGTAAGTCTATATGACTTCAGAAA
TGTTAAATAGACTAACCTCTAACAACAAATTAAGAGTGAATGTTTCAAGGTGAT
GCAATTATTGATGACCTATTTTATTTTCTATAATGATCATATATTACCTTTGTAAT
AAAACATTTTCCC

Sequence ID - 576

nt: 541

CAGCCAACCCAGAAGGAGCCAGTCTACAACCTATGCCTGATCCTCCTCATGGCAGG
CCACGAAGCATTGCTGCCATGTGTTGAATTATAAAACCCACATTGCTTTTTGAACC
CTGTTGCGGGTAAAAATAACCAAATTATCAGTCCTTGGAAACCCAGGCAATCAAG
TGAGTACAAGGTAAAGATAAGTATGGTTTAGAGGAGAAATTATGTTCTGAACTG

GTGTCCTTTGATGGCAGCGTCAGCCTTGCTAAGTCAGAGTAGAGGGAGCAGTGAC
CTTAATAAGCTTTGGTGAGCATCATGTGCACGCGTGGGTGGGAGTCCCTTTCACT
GATGCTTTTAAAAGTGCTTTTGCAGACCCTGGAAGGGATCCTCCACACATATGAG
GTGTGGGACAGGTAGGCCAGAGAGGATTAGCCCTGCTTTTCGAGACTAGAAATCT
ACAGTCCTGAAGGAGCAGTAATTAATTGGTACACCTGTCAGGGCCAGCCCCCAGG
TCTCCTGGCTTTTTCCAGGTTTTCTGTCTCACATGATTTTGCTTTT

Sequence ID - 601

nt: 522

TCGACCGGGTTTGGAGCAGTGCCTTGTTTGCTGTGCAGCGGATACTCTACAGGTA
CATTTCCCTTTTTGGAACCAAAGGGAGGGATTGACAATATTGATGGTAGATCTT
TTTTCTTTAGCAAGAATTAAGGATTTTGGTGGGTGGGGGGAGGCTTCTGTGGGGA
CCAAGACAATGTACTGTCAGTCAGGATTTAAGTCGAACCTACCTCATCCCTTGCCC
CAGAGAACAGTTGATCGTGTTTTAAACCAAAGGTGCGGAATGGAGAGAGGGAG
GCGGTGCATTGCAGCTTCCGATAGAGCTTTTTATTTTGGATATCAGGAACCAATT
TTGAAGATTTCTTAAGAAAGTCATTTACATCAGGGACATGAAGAGCAAAGTAGGT
ATTTTGGTCAGTACTTGAATTTGATAGGCTTTATGCAAACAACCTCTCCCTCTGCT
GGAGTCTGGCAAGTTTGCTTTTCACTGGACGCTAATTCAAGTGCCATACAAAAC
AAAATAANAGTTTTACTTATAACACA

Sequence ID - 612

nt: 576

GAGAAATATAAGATTATGTATAGATCAAATCTACCTCTATTTGGTGTCTGAAAG
AGATGAGGAGAATGGGACAAACTTGGAAAGCTTATTTCAAGATAACATTCCTGA
GAACTTCCCCAATCTTGCTAGAGAGGCCAACATTAAAATTCAGTAAATGCTGAAA
ACTCCAGTAAGATATTTCTTAAGAAAATTATTCCCAAGATATATACTCATCAAATT
ATCTAAGGTCAAATGAAGGAAAAAATTTTATAGGCAGCTAGAGAGAAATGTCAG
GTCACCTACAAAGAGAATGGCATAAGACAAAAAGTAGAACTCCCAGCAGAAACT
CTAAAAGCCAGAAGAGATTAGGGGGCCAATATTTAACATTCTGAAAGAAATTCCA
ACAAGGAATTTTCATATCCAGCCAAACTAAGCTTCATAATTGAAGGAGAAATAAG
ATATTTTCCAGACAAGCAAATGCTGATGAAATCCATCACCACCAGACCTGCCTTA
TAAGAGCTCCTGAGGGAAGCACTAAATATTGAAAGGGAAGAACTTTATGAACCA
TTTCAAAAACACATTTAAGTNCACAAAGCAG

Sequence ID - 628

nt: 419

AAGAGAAAGGACTCAGTGTGTGATCCGGTTTCTTTTTGCTCGCCCCTGTTTTTGT
AGAATCTCTTCATGCTTGACATACCTACCAGTATTATTCCCGACGACACATATACA
TATGAGAATATACCTTATTTATTTTTGTGTAGGTGTCTGCCTTCACAAATGTCATT
GTCTACTCCTAGAAGAACCAAATACCTCAATTTTTGTTTTGAGTACTGTACTATC
CTGTAAATATATCTTAAGCAGGTTTGTTTTCAGCACTGATGGAAAATACCAGTGTT
GGGTTTTTTTTAGTTGCCAACAGTTGTATGTTTGCTGATTATTTATGACCTGAAA
TAATATATTTCTTCTTCTAAGAAGACATTTTGTTACATAAGGATGACTTTTTTATA
CAATGGGAATAAATTATGGCATTTTTT

Sequence ID - 635

nt: 592

TGAGCGTTGGGCTGTAGGTGCTGTGCTGTGTGATCCCCCAGAGCCATGCCCGAG
ATAGTGGATACCTGTTTCGTTGGCCTCTCCGGCTTCCGTCTGCCGGACCAAGCACCT
GCACCTGCGCTGCAGCGTCGACTTTACTCGCCGGACGCTGACCGGGACTGCTGCT
CTCACGGTCCAGTCTCAGGAGGACAATCTGCGCAGCCTGGTTTTGGATACAAAGG
ACCTTACAATAGAAAAAGTAGTGATCAATGGACAAGAAGTCAAATATGCTCTTG
GAGAAAGACAAAGTTACAAGGGATCGCCAATGGAAATCTCTCTTCTATCGCTTT
GAGCAAAAATCAAGAAATTGTTATAGAAATTTCTTTTGAGACCTCTCCAAAATCT

TCTGCTCTCCAGTGGCTCACTCCTGAACAGACTTCTGGGAAGGAACACCCATATC
TCTTTAGTCAGTGCCAGGCCATCCACTGCAGAGCAATCCTTCCTTGTCAGGACACT
CCTTCTGNGAAATTAACCTATACTGCAGAGGTGTCTGTCCCTAAAGAACTGGTGG
CACTTATGAGTGCTATTCGTGATGGAGAAACACCTGACCCA

Sequence ID - 636

nt: 572

CTTANAAGAGTTGCTCATTACACCCACGCCCTTGCCCAAGGCTGGCCCACTCAG
AGCGAAACTTAACTTTTGTCTGGATGGGAAGAGAAGTAAGTCTACCCCGAGGTG
CCATGTTGAAGAGTGAGAGGTCCAAGTGATTCTGTGCATTGAAACCAAGACACCC
CACCCAGAACACTTCTTCCCTCCCTCAGCCCAAACCAAAGGCTGGGGTTCTCATC
TCCAAGTGGCTGTTCTCCAACCTTCCCAAGCCGCTTGCATTCCCCAGACTGGACTA
CTGTGGCGGTTAGGTTAGATTTGAAGACGGGGGCCAGGCTGGGTATGAACGGGT
GCAGCCCTCTTCTCCTCTTCCCCCCCCACATCTCTCATGAGAGAGGTAGTGGCATT
CCTTCTCAGGGAGCTTCAATGGGAAAGGTCTCGAAAGCTTCAGGAGGAGCAGAA
TACCAACGCAGGGGGATGGCTGTAACGATCTCACCGTCTCCTAACCTCAGTCCCT
TTTTTGAGAGTGAATGGTGGAGGGTGGGAAAGGGACCCAAATTTGTAGATCTCTT
TGCTGGGGGAGGGGAANGATG

Sequence ID - 637

nt: 482

TTAAACAGGCGCAGGGGTAAAAATGAGAATGAATCTGAAAAAAGAGAGTTGGT
GTTTAAAGAGGATGGACAAGAGTATGCTCAGGTAATCAAATGTTGGGAAATGG
ACGATTGGAAGCATTGTGTTTTGATGCTGTAAAGAGGTTATGCCATATCAGAGGG
AAATTGAGAAAAAAGGTTTGGATAAATACATCAGACATTATATTGGTTGGTCTAC
GGGACTATCAGGATAACAAAGCTGATGTAATTTTAAAGTACAATGCAGATGAAG
CTAGAAGCCTGAAGGCATATGGCGAGCTTCCAGAACATGCTAAAATCAATGAAA
CAGACACATTTGGTCTGGAGATGATGATGAAATCCAGTTTGACGATATTGGAGA
TGATGATGAAGACATTGATGATATCTAAATTGAACCAAGTGTTTTTACATGACAA
GTTCTCTGAGGATGGTTCTACAGTTGGGATTTTGGCCATCATCAAC

Sequence ID - 639

nt: 624

GACACACGAGCATATTTACCTCCGCTACCATAATCATCGCTATCCCCACCGGCG
TCAAAGTATTTAGCTGACTCGCCACACTCCACGGAAGCAATATGAAATGATCTGC
TGCAGTGCTCTGAGCCCTAGGATTTCATCTTTCTTTTACCGTAGGTGGCCTGACTG
GCATTGTATTAGCAAACCTCATCACTAGACATCGTACTACACGACACGTACTACGT
TGTAGCCCACTTCCACTATGTCTATCAATAGGAGCTGTATTTGCCATCATAGGAG
GCTTCATTCACTGATTTCCCTATTCTCAGGCTACACCCTAGACCAAACCTACGCC
AAAATCCATTTCACTATCATATTCATCGGCGTAAATCTAACTTTCTTCCACAACA
CTTTCTCGGCCTATCCGGAATGCCCGACGTTACTCGGACTACCCCGATGCATAC
ACCACATGAAACATCCTATCATCTGTAGGCTCATTCAATTTCTTAACAGCAGTAAT
ATTAATAATTTTCATGATTTGAGAAGCCTTCGCTTCGAAGCGAAAAGTCCTAATA
GTAGAAGAACCCTCCATAAACCTGGAGTGACTATATGGATGCCCCCACCCTACC
ACACATTCGAAGAA

Sequence ID - 649

nt: 425

CAAAAAACGAAGAAAAGTGACGACAGTCTGAGGGACTTATGGGAGATCATCAA
GTGAACCACTATATGTGTAATGTAAGTCTTGGAATGAGAAGAGAGAAGGAGAAG
GAGGAGAGAGCTTATTTGTAGAAATAATGGCTGAAAACATCCCAAACCTTTCTTT
TTTTGAGGAAAGAAATAGGCATACAAGTTCAAGAAACTCAAGGAAGTCCAGAGA
GGACAATTCTAAAGACACCCCCTCTAACATACATTATAATCAAATTGTCAAAAGT

AAAATACAAAGAGAATCTTTTAAATTGACAAGAGAAAAGCAGCTGGTCACGTTCAAGGGAGTTCTATAAGAATTTTCAGCAGATTTCTCAGCAGAAACCTTGCAGGCCAACAGGCAGTGGGATGATACATTCAAAGTGCAAAAAAAAAAAAAAAAAA

Sequence ID - 651 nt: 251

CTTTGGGAGGCCGAGGCGGGCGGATCACTTGAGGTGAGGGGTTGAGACCAAGTCTGGCCAACATGGTGAAACCCCACTCTACTAAAAATACAAAAGTTAGCCAAGTGTGGTGGCAAGTGCCTGTAATCCCAGCTACTCGGGAGGCTGAGACAGGAGAATCACTTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCCAAGATCGTGCCACTGCACTTCACCTGGGCAACAGAGCAAGATTCCGTCCATCTC

Sequence ID - 665 nt: 345

ACCGGCGACATGGCCAAACGTACCAAGAAAGTCGGGATCGTCGGTAAATACGGGACCCGCTATGGGGCCTCCCTCCGGAAAATGGTGAAGAAAATTGAAATCAGCCAGCACGCCAAGTACACTTGCTCTTTCTGTGGCAAAACCAAGATGAAGAGACGAGCTGTGGGGATCTGGCACTGTGGTTCCTGCATGAAGACAGTGGCTGGCGGTGCCTGGACGTACAATACCACTTCCGCTGTACCGGTAAAGTCCGCCATCAGAAGACTGAAGGAGTTGAAAGACCAGTAGACGCTCCTCTACTCTTTGAGACATCACTGGCCTATAATAAATGGGTTAATTTATGTA

Sequence ID - 675 nt: 591

GTATAGAAAATAATGTCCCCAGNGCATAGAAAAAATGAGTCTCTGGGCCAGTGAATACAAAACATCATGTGCGAGAATCATTGGAAGATATACAGAGTTTCGTATTTTCAGCTTTGTTTATCCTTCCTGTAAAGAGCCTCTGAGTTTTTAGTTTTTAAAAGGATGAAAACTTATGCAACATGCTCAGCAGGAGCTTCATCAACGATATATGTCAGATCTAAAGGTATATTTTCATTCTGTAATTATGTTACATAAAAGCAATGTAAATCAGAATAAATATGTTAGACCAGAATAAAATTAATTATATTCTGGTCTTCAAAGGACACACAGAACAGATATCAGCAGAATCACTTAATACTTCATAGAACAAAAATCACTCAAAACCTGTTATAACCAAAGAATTCATGAAAAAGAAAGCCTTTGCCATTTGTCTTAGAAAGTTATTTTTTAAAAAAAATCATACTTACTATTAGTATCTATGGAAGTATATGTAACAATTTTTATGTAAAGGTCATCTTTCTGTGATAGTGAAAAAATATGTCTTTACTAAGTTGAAATGAATACTTTCTGNCTTTGCTAATGGATAGTTATT

Sequence ID - 684 nt: 545

GTGGAAGNGACATCGTCTTTAAACCCTGCGTGGCAATCCCTGACGCACCGCCGTGATGCCCANNGAAGACAGGGCGACCTGGAAGTCCAACCTACTTCCCTTAAGATCATCCAACTATTGGATGATTATCCGAAATGTTTCATTGTGGGAGCAGACAATGTGGGCTCCAAGCAGATGCAGCAGATCCGCATGTCCCTTCNCGGGAAGGCTGTGGTGCTGATGGCAAGAACACCATGATGCGCAAGGCCATCCGAGGGCACCTGGAAAACAACCCAGCTCTGGAGAACTGCTGCCTCATATCCGGGGGAATGTGGGCTTTGTGTTACCAAGGAGGACCTCACTGANATCAGGGACATGTTGCTGGCCAATAAGGTGCCAGCTGCTGCCCCGTGCTGGTGCCATTGCCCCATGTGAAGTCACTGTGCCAGCCCAGAACACTGGTCTCGGGCCCCGATAAGACCTCCTTTTTCCAGGCTTTAGGTATCACCCTAAAACTCCAGGGGCAACCATTGAAATCCTGAGTGATGTGCACTGATCAAGACTGG

Sequence ID - 688 nt: 569

CTTTAGCCAGCCTGATCAGAAAAAACAAGAGGAAAGACGTTAGATTACC AACATCAAGAATGTGAGTTATGATATCACTACAGACTCTCCAGGTATTAAGCA TAATTAGAGAATGATATGAGCAGCTATATGCAAATAAGTTCAACATTGGACAAATGGACAAATTTCTTGAAAGATAAATTATGAAATTTTATTCTGAAAGAACTACATGA

CCTTAATTGTCTTACATCTATTAAATAAGTGGAAATTGTAGTTTAGAAACTTTCCC
ACAAAGAAAACCTCTAGGCCAGATGGCATCAAATAATATTCAGATGAATGAAA
TGGAGAAAGGATAGCCTTTTCAACAAATGGTGGTGGAAACAATTGGATTTCCATAT
GCAAAAAAATAGAGATGGACGCAGAGGTGTGTGCTTAGGAGGCTGAGGTGAGAG
GATTGTTTGAGGCCAGCCTGGGCAACATAGCAAGACCCCATTTCAAAAACAAAA
ATAAAGAACTTGTAGCCTTACCTTGTGCCATATTATGAAAATGTATCATAGGCTT
AAATGTGAAACGTAAAACAAAA

Sequence ID - 701

nt: 579

CTTTGGAGCTTCTGTCTGTGCTGTGGACCTCAATGCAGATGGCTTCTCAGATCTGC
TCGTGGGAGCACCCATGCAGAGCACCATCAGAGAGGAAGGAAGAGTGTGTGT
ACATCAACTCTGGCTCGGGAGCAGTAATGAATGCAATGGAAACAAACCTCGTTG
GAAGTGACAAATATGCTGCAAGATTTGGGGAATCTATAGTTAATCTTGGCGACAT
TGACAATGATGGCTTTGAAGGTAATTAATAATTATCAAATTGGTGCTTGATTTCTGC
TTTTAAAATGGTTTATGGAAGAAAATATGATTAAAGTTTTGTATTGTTTTCTTCC
TATAGAAGATGGAGCCAGAATGGCATGCTAAGTTTTTCTTTTCTTTAGTGTTATA
TATGACTTCTCCTCAATTGTACCCATTGATCTTTACCACTGTTAATAATGGATGA
TATTCAAATACCTTATTTCAAGTATTCTAAGGCACCATTGATTAGAACTGCATT
ATTATTTATGTGTCCCTAAAAGCTACCTATTAAGCTGTTACACCACCATTTTTCT
GTTAAGAAAATCCTGATTTAGAA

Sequence ID - 726

nt: 260

CGGGGTCTGTACCGGGCTGGCCTGTGCCTATCACCTCTTATGCACACCTCCCACCC
CCTGTATTCCCACCCCTGGACTGGTGGCCCCTGCCTTGGGGAAGGTCTCCCCATGT
GCCTGCACCAGGAGACAGACAGAGAAGGCAGCAGGCGGCCTTTGTTGCTCAGCA
AGGGGCTCTGCCCTCCCTCCTTCTTCTTCTCATAGCCCCGGTGTGCGGTGC
ATACACCCCCACCTCCTGCAATAAAATAGTAGCATCGG

Sequence ID - 736

nt: 641

GGAATTCCAAGTGCTTGGGGATAATGATACCTCTGACCTTTCTTCCTTTTGGGAAG
TACTTGAGTGTGCAGCTGCATGAGGCCTCAGCAGGAGAGAGATTTTAGGTCCAAG
AAGCTATACCAGTAGGACAAGGCAGGAAAATACTACACTTTCAGGATCAAGCCC
CTCTGACTCTCATTTGGAACTGGATGTTTGCTAAGCACCTGCTTCTTAAGGATGC
CGAGGGATTTAATGATACTCCCAGAAACCTGGAGAGATTAATGGGGCCTATGGA
GAAGTGCTCTGAACTCAGTGTTGGGACTTGAATAAAATTAACCATTGTCATGTTTT
CAGAACAATAAGCTGTTTTATATTTTATGTGCATGAAAGCCCTAGAACTAAGTT
GTGTTATTTCCAGAAATGAAATAGATCCACAGTTAGATGATGTGGCCATTAGGA
AGTACCAAATTTATAAAAATCACTGGAGGTCTGTCTGAGCAGTACCTAATAAAAT
ATAGTATACTGAAAGTGAACAGATACTTTGTCTCTTTCTTTGGCTGCTTGATCTTT
ATCTGTGTCTGCCGTACAGTGCACCCTTAAAGTATTCTACACCAGTGCTTCTCAA
CTGGAAATGTGCATGTAAGTCACCCANGGGTCT

Sequence ID - 757

nt: 583

GAACCTGCGGAGGGACTTCAATCACATCAATGTAGAACTCAGCCTTCTTGGA
GAAAAAAGAGGGCTCCGGGTTGACAAATGGTGGGGTAACAGAAAGGAACTGG
CTACCGTTCCGACTATTTGTAGTCATGTACAGAACATGATCAAGGGTGTACACT
GGGCTTCCGTTACAAGATGAGGTCTGTGTATGCTCACTTCCCCATCAACGTTGTTA
TCCAGGAGAATGGGTCTCTTGTGAAATCCGAAATTTCTTGGGTGAAAAATACAT
CCGCAGGGTTCCGATGAGACCAGGTGTTGCTTGTTCAGTATCTCAAGCCCAGAAA
GATGAATTAATCCTTGAAGGAAATGACATTGAGCTTGTTTCAAATTCAGCGGCTT

TGATTTCAGCAAGCCACAACAGTTAAAAACAAGGATATCAGGAAATTTTTGGATG
GTATCTATGTCTCTGAAAAAGGAACTGTTTCAGCAGGCTGATGAATAAGATCTAAG
AGTTACCTGGCTACAGAAAGAAGATGCCAGATGACACTTAAGACCTACTTGTGAT
ATTTAAATGATGCAATAAAAGACCTATTGATTGG

Sequence ID - 758

nt: 424

CTTGGCTCCTGTGGAGGCCTGCTGGGAACGGGACTTCTAAAAGGAACTATGTCTG
GAAGGCTGTGGTCCAAGGCCATTTTTGCTGGCTATAAGCGGGGTCTCCGGAACCA
AAGGGAGCACACAGCTCTTCTTAAAATTGAAGGTGTTTACGCCCGAGATGAAACA
GAATTCTATTTGGGCAAGAGATGCGCTTATGTATATAAAGCAAAGAACAACACA
GTCACTCCTGGCGGCAAACCAAACAAAACCCAGAGTCATCTGGGGAAAAGTAACT
CGGGCCCATGGAAACAGTGGCATGGTTTCGTGCCAAATTCCGAAGCAATCTTCCTG
CTAAGGCCATTGGACACAGAATCCGAGTGATGCTGTACCCCTCAAGGATTTAAAC
TAACGAAAAATCAATAAATAAATGTGGATTGTGCTCTTGT

Sequence ID - 764

nt: 626

GATTTTTTTTTTTTTTTTGGAGATGGAGTCTTTCTCTGTGCGCCAGGCTGGAGTGCAG
TGGTGAAATCTCGACTCACTGCAACCTCCGTCTCCTGGGTTCAGCAATTCTCCTG
CCTCAGCCTCCTGAGTAGCTGGGATTACAGGCACCAGCCACCACGCCCGGCTAAT
TTTTGTATTTTAGTAGAGACAGGTTTTTACCATTGTTGGCTAGGCTGATTTTGAAC
TCATGACCCCAAGTGATCTGCCCGCCTCGGCCTCCCAAAGTGCTGGAATTACAGG
TGTGAGCTACCACTCCCAGCCAATGATTACATTTATAAGGTAAAATAACTTGTGC
CAATCTGTACAAGTGAATTCAGATTTAAAATTTTAATTGTAAAAAGATATCCAGG
TGATATTTCTCCCTGAATAATTTAGTTTCTTTTCTATTTCTTGATATAAAAGTACT
CAGCATTGAAGTAATTGCTATCTTCACATTTCTTCCTATTTGAGCTGTCTAAATAA
GTAGTCCTACATATTTTCCCCCAACACAAAAAACCCAGAAAAGAATTATTTTAT
ACTGGATTTTTTTGGTTGTAGCAGGAACCTAAAGNGCCAATTGTAACATGCATG
TTCTTTTTGGCAAA

Sequence ID - 785

nt: 556

CTTTTCTCTGGGTATAGATTTACCCTAGCACCTATCTCATTATATTGAATTTTCCAG
CATATTTAAATAAACTATTAATTAGTCACACTATTTCTTAAAAGTCACACTATCAA
CTAATCGTGACCGCAATTATCTAGGGGTGATAATCTGCTGAGTCTACTCTTTAAAT
ACACTGGGACCCAGCATATTGAGTTATATTGGCACAGAACTTCACTCTGGGTAT
AGATTTACCCTAGTACCTTGCCGGCAGGATCCTATTATTCATGGTTGTACAAGCA
AGGTTACAGGAAGAGGCTGGCACAGAGAAGGTACCTGGTAACTGTTGTTTGAGG
CTGAATTCAGCTCAACTCAGCTCCAGTAGAGATGGTGTCCCTTCTCTACCGTGT
GAGATAGTGTGCAGTCCCTTCTTAAGGGCTGTTACCCACCGCAATAGGACTTCTC
AGCTTCAACTTTTAAATTTCTCTGCTCCCGCTGGGACCCACCCGCTTCAAAAATCA
TCATGGNGGNTTTAGCACCAATTTAGTAAACACAACTGTCTGAAATATTTTGA
T

Sequence ID - 808

nt: 641

CCGGGTTTTAGTATTTAACCAAGAGCCTTTTAAATATTGAAAACCCATAGTTCAG
AAAATGTTAGTATTGCTGCCCTTCTTCACATAAATTTTTTTTAAATTATACTATTA
TTTTGCTTAATTTTATATTGGGTTAAAACAACCTTCAAGAAGGTTAACTAGGAAA
GAAGACCTTTTTGTTTTATTTTTACTATTTATATATAGAAGACAAATCAGCATTTG
GTGATAGTTTTACATGACCAGTTATCAAACGGTCATAGTATGAAGTGTGCAGTTG
TTCATTATTAGTAAATTATGTTTGATTTTTTAACTATTTAGTACTAATAGTTGAGA
TGAAAACCTGAAGAAAAATGCCAATGTGACGTTTGTGTATAGCTAGCCTTAAAAAA

CTTCCCATGTTTTTAGGTGACTTTTTTCCCCCTCTTAGTACTCTGGAGAAACAATG
AAGATGGGCCATCTCAATTCCAGATGTAAACAAAAAGTAATTTTTATTCAACAT
TTAATGTAACGTCTATTATTGNGGATTCTTGNCTTGNGTATTTTCTTTCCCTTATTC
AAGTAATATAGAATAACTTTCCTTAAAATGATTTGATCCAAGATACGTCATTTCTG
TATTGGCAAAATGCCNCTATTAAAGTGT

Sequence ID - 814 nt: 132
GTAAAGTGATACATTTTTTATACCAAATGTGTTTATTTTTTTGTGCAAGTAATCCT
TAAATTGCAATTGTATTAGGTGTTAAATAAAGTTTTTAAAAAATTAAAAAAA
AAAAAAAAAAAAAAAAAAAAA

Sequence ID - 821 nt: 370
AAAGAGCTCCCAAATGCTATATCTATTTCAGGGGCTCTCAAGAACAATGGAATATC
ATCCTGATTANAAAAATTTGGATGAAGATGGATATACTCAATTACACTTCGACTC
TCAAAGCAATACCAGGATAGCTGTTGTTTCANAGAAAGGATCGTGTGCTGCATCT
CCTCCTTGGCGCCTCATTTGCTGTAAATTTGGGAATCCTATGCTTGGTAATACTGGT
GATAGCTGTGGTCTTGGGTACCATGGCTGGTTTCAAAGCTGTGGAATTCAAAGGA
TAAATTAATGAAGAAAACAAGCGGAGCTGAAGAAGAAAGTACAATATGGTGCTG
TCTTCTAATGAAATAAATTCATAAATGGACATTAATAA

Sequence ID - 837 nt: 603
TGAGGNTGGTCATGATGCANAAGCTACTCAAATGCAGTCCGGCTTGTCCTGGCTCT
TGCCCTCATCCTGGTTCCTGGAATCCTCAGTTCAAGGTTATCCTACGCGGAGAGCC
AGGTACCAATGGGTGCGCTGCAATCCAGACAGTAATTCTGCAAACCTGCCTTGAAG
AAAAAGGACCAATGTTTGAAGTACTTCCAGGTGAATCCAACAAGATCCCCCGTCT
GAGGACTGACCTTTTTTCCAAAGACGAGAATCCAGGACTTGAATCGTATCTTCCCA
CTTTCTGAGGACTACTCTGGATCAGGCTTCGGCTCCGGCTCCGGCTCTGGATCAG
GATCTGGGAGTGGCTTCCCTAACGGAAATGGAACAGGATTACCAACTAGTAGACG
AAAGTGATGCTTTCCATGACAACCTTAGGTCTCTTGACAGGAATCTGCCCTCAGA
CAGCCAGGACTTGGGTCAACATGGATTAGAAGAGGATTTTATGTTATAAAAGAG
GATTTTCCACCTTGACACCAGGCAATGTAGTTAGCATATTTTATGTACCATGGNT
ATATGATTAATCTTGGGACAAAGAATTTTATAGAAATTTTAAACATCTGAAAA

Sequence ID - 839 nt: 71
ATTTATCTAATATTTGGTTTAATAAAATGTGAATAATGAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAA

Sequence 849 nt: 622
TTTTTTTTTATTTTTTGAGAATGGAGTCTTGCTCTGCCGTCCAGGCTAGAGTTCAGT
GGTGCGATCTCAGCTCACTGCCACCTCACCTCCTAGGTTCCAGAGATTCTTGTGCT
TCAGCCTCCTCAGTAGTTGAGAATACAGGAACACGCCACCACGCCTAGCTAATTT
TTGTATTTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACCTC
CTGGCCTAAGTGACCCACCTGCCTCAGCCTCCCAAAGTGCTGGGATTATAGGCGT
GAGTCATTGTCCCCAGCCGGATGTTTTCATCTTGATTGCCTTAGTTTCTAAATCT
CATCCTCTCCATTTTCTCCTGTAGTAGTCACAGAGAACCAAATTCTGTCAAGTTA
TGAACTAAAGTCTCTCTTCCACAAGTCTTCCGTGTGTTCTGCCTCAAGTGAAGTTG
AAAGAACATCAGTTTGTGGGAAGGTTGAAGACCGAATGATCTGCTGGGAAATCA
CTGAGGCATTGCCATTCTCTTGAGGAATTTCAATTTTCATCGAAGTTTCGGTTTATA
TCCCTTTCTTGGTGAGTACTATTGCTGTTATGTAAATTAATGAGTCGTCATCCTT
CTTNTGAGC

Sequence ID - 860

nt: 501

GTGAAATCACTTTCATGGATTATTAATGGATTTAAGAGGGGCATCAATCAGCTCAA
CTCAAGATTTTCATAATCATTTTTAGTATTTAGATTGTGCCTCAAAGTTGTAGTACC
TCACAATACCTCCACTGGTTTCCTGTTGTAAAAACCTTCAGTGAGTTTGACCATTG
TGCTCTTGGCTCTTGGGCTGGAGTACCGTGGTGAGGGAGTAAACACTAGAAGTCT
TTAGTACAAAACCTGCTCTAGGGACACCTGGTGATTCTACACAAGTGATGTTTAT
ATTTCTCATAAAGAGTCTTCCCTATCCCAAGGTCTTCATGATGCCAGTAGCCATAT
ATGATAAATTATGTTTCAGTGATAACTTAGTTATCAGAAATCAGCTCAGTGGTCTTC
CCCGCCATGATTCACATTTGATGAGTTTTTAAAAATCAAAGTGATTTTGAAAATCT
CTAATGGCTCAGAAAATAAAAACATCCAGTTTGTGGATGACTATATTTAGATTTC
T

Sequence ID - 891

nt: 626

GGCAGAGGTTGCAGTGAACCTGAGATCATGCCATTGCAATCCAGCCTGGGCAACA
NGAGTGAGACTCCATCTCAAAAAAAAAAAAAAAAAAGACAAGAGTNTCCACTCTA
AACACTTNTATTCAACATAGTCCTGAAAGTCGTAGCCACAGCAATTTAACAAGAT
AAAGCAATAAAATGTATTCAAATAGAAAAAGAGGAAGTCAAATTATCTTCACTG
GNGATATAATTCTCTACCTGGGAACTTCACCGAAAAAGATTTACCAAAAAGATT
TCTAAGCCTAAATAATGACTTCAGCAAAGTCTCACCATACAAAATCAACATACAC
AAATGAGTAGCATTCTGTGCACCAATAATATTCAAGCTGAGAAAAAAGAACA
TGGTTCTATTTACAATAGCTACAAACAAAAAATATGTACCTAGTAATACATTAA
ATCAAGGNGGTAAAATATCTNTACAACAAGAACTACAAAACCTGCTGAAAAAAA
TAGAGACACGCAAATAAGTAAAAAGGCACTCCATGCTCATGAATTTAAAGAATC
AATATAATTAAATGTCCGNGCTGCCTAAAGCAACTTACAGATTAAAGGCTATTT
CTCTCAAACCTATAAATGCACCTTTTA

Sequence ID - 897

nt: 509

GCAAATCTACACATTTGATTAAATGATAGGGAACCTATGCACACACATAATACATA
TAATGCTAGTTTCTTGGTTTTGATATTGTACCATAGTTATGTAAGATGTAACCATT
GGGGGAAACTGGGTGAAGGCTACATGAGACCTCTCTGTACTTAATCTTTGCAACT
TATGTGAATCTATAATTATTCCAAAATAAAAAAGTTTTAAAGAACCTAAGTATCCT
TATTACTGAGGGTCATCGTGCTAGACAGCAAGGTTGGGCCAGAGCTTCTAGTTAT
TTAAAATACTAAATACCAGCCTGGGCAACATAGCAAGACCCTGCCTCTACAAAAA
GCAAAAAAATTAGCTGGGCATGGTGGTACATGCCTGTGGTCCTAGTTACTCTTGG
AGGAGTCTGAGGTGGGGAGCTTGAGCCTAGGAGTTTGAGGCCGCAGTGAGCCTT
GATTGTGTCTCTGTACTCCAGTCTGGGCCACAGAGCAAGACCCGGTCTCTAAAAA
TAAATAAATAAATA

Sequence ID - 939

nt: 513

GGAACCCAGTGTATTACCTGCTGGAACCAAGGAACTAACAATGTAGGTTACTAG
TGAATACCCCAATGGTTTCTCCAATTATGCCCATGCCACCAAAACAATAAAACAA
AATTCTCTAACACTGCAAAGAGTGAGCCATGCCTGTAAACACTGTAAAGAATGTA
ACATGTGGGGGACACACAGGGGCAGATGGGATGGTTTATGTTTAGGATTTTATTAG
TGCATGCCCTACCTCTGGGGGAACGTCCCATCTGAGGTTTTCTTCTCGGTGGGGG
GATTTAACTTCTGTCTAGGGAAAAACAGTGTCTGATGAGGAGTGTTCACACACA
GGCTACATGAATTCCTTATACCAGTGCAGAAAGCAGCCAGGAGTCCCCGTTGGAA
AAGAACAATGCCACTCTCTTTTATGTATCTTGGTTCTGCAACTCATTTGTTGTAAG

TAGGGTTAATCGAGTATCAGGTTACAGTATCCTGCCCTTATTATTTTATGATTCA
CTGACTCAAGTTCCA

Sequence ID - 1056

nt: 435

TCGCTTGTAAGCCTGAGACAGCTGCCTGTGTGGGACTGAGATGCAGGATTTCTT
CACACCTCTCCTTTGTGACTTCAAGAGCCTCTGGCATCTCTTTCTGCAAAGGCATC
TGAATGTGTCTGCGTTCCTGTTAGCATAATGTGAGGAGGTGGAGAGACAGCCAC
CCCCGTGTCCACCGTGACCCCTGTCCCCACACTGACCTGTGTTCCCTCCCCGATCA
TCTTTCCTGTTCCAGAGAAGTGGGCTGGATGTCTCCATCTCTGTCTCAACTTCATG
GTGCGCTGAGCTGCAACTTCTTACTTCCCTAATGAAGTTAAGAACCCTGAATATAA
ATTTGTTTTCTCAAATATTTGCTATGAAGGGTTGATGGATTAATTAATAAGTCAA
TTCCTGGAAGTTGAGAGAGCAAATAAAGACCTGAGAACCTTCCAGA

Sequence ID - 1083

nt: 198

GCGCGTCGACTTTGTTTAGACATTGAATGACTTTGTTAAAGGCACAATTAATCAC
ATTGGTTGTACTCTGNNGACAGCCTTCTTTAAAAAAAAAAATAAACAATTTAAAC
AA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANTTTTAACC

Sequence ID - 1084

nt: 198

GCGCGTCGACTTTGTTTAGACATTGAATGACTTTGTTAAAGGCACAATTAATCAC
ATTGGTTGTACTCTGNNGACAGCCTTCTTTAAAAAAAAAAATAAACAATTTAAAC
AA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANTTTTAACC

Sequence ID - 1074

nt: 689

GGGAGGCGGAGGCTGCAGTGAGCTGAGATCGTGCCACTTCATTCCAGCCTGGGC
AACAAAGCGAACTCTGTCTCAAAAAAAAAAAAAAAAAAAATTTGTTGACTGTT
GTAATTTAAAGCTTGTCAATTTTTATTAGTAATAACACTCATTAGTGTAGTATCT
ATGATGAACCAGGTTCTGCACAAAGTACCTTATGTTTCATGGCCTCATATCGTCTC
TCCAAAACCTCTGCAAGATAGGATTCATCACCCTTATAGGGAGAGATCTGAAAGT
TTAAATTTGTACCCAAGGTCACACAGCTGGTAAGTGCCAGAGCTGGGATTCCGTA
GGGTGTTCANAGTGCTCTCCTGCCGTAGGCTTATCACAAAAGTCAAAGTTTGG
TCATAATAAGCCTGAAGTTTGGCAGGATTTAAAAATAGTCACCANACTTTTGAG
TTGGAGCATCCCACCTCACTGCTGTTACCTTCTGTGGCAGGGAGAGTCATCATT
CCATTTAGCTTGTGGAATATCTTGTCAATTAACATTCTCATGCAAAAGCCATTTTA
TGGTGCCCAATGAANATGGTTAAGCTACTGCCCAAGCCTNTGGAAGCCTTCCTA
ATTTTGGACTTGCACTATGCAAATTGNATAATATTTTCTCTACCCTAAGCCAAATA
TTTTCTTCACTTTTCATTCTAC

Sequence ID - 1099

nt: 561

TGCATGCTTGTGGATTGGAAAACTTTGGAGACTGATTACTTTTCATTATATATGT
GTCACAGTGAAACAGCTTTTATGTGTCATGTAAGATTACTGCTTGCCTCTCTAAGG
AAGGTCGTGACTGTTTAAATAGACGGGCAAGGTGGAACCTTTTGAAAGATGAGCT
TTTGAATATAAGTTGTCTGCTAGATCATGGTTTGTATTGAACTAACAAGGTTTGCA
GATCTGCTGACTTATATAAAGCTTTTTGATTCCCTACTAAGCTTTAAGATTTAAAAA
ATGTTCAATGTTGAAATTTCTGTGGGGCTCTATTTTGTCTTTGGCTTTCTGGTGAG
AGAGTGAGGAAGCATTCTTCCCTCACTAAGTTTGTCTTTCTTGTCTTCTGGATAG
ATTGATTTTAAGAGACTAAGGGAATTTACAACTAAAGATTTTAGTCATCTGGTG
GAAAAGGAGACTTTAAGATTGTTTAGGGCTGGGCGGGGTGACTCACATCTGTAAT

CCCAGCACTTTGGGAGGCCAAGGCAGGCAGAACACTTGAAGGAGTTCAAGACCA
GCGTGG

Sequence ID - 1139

nt: 503

CAGCACTGCCAGTGGAGATGGGCGTCACTACTGCTACCCTCATTTACCTGCGCT
GTGGACACTGAGAACATCCGCCGTGTGTTCAACGACTGCCGTGACATCATTCAGC
GCATGCACCTTCGTCAGTACGAGCTGCTCTAAGAAGGGAACCCCCAAATTTAATT
AAAGCCTTAAGCACAATTAATTAAGTGAAACGTAATTGTACAAGCAGTTAATC
ACCCACCATAGGGCATGATTAACAAAGCAACCTTCCCTTCCCCCGAGTGATTTT
GCGAAACCCCTTTTCCCTTCAGCTTGCTTAGATGTTCCAAATTTAGAAAGCTTAA
GGCGGCCTACAGAAAAAGGAAAAAAGGCCACAAAAGTTCCCTCTCACTTTCAGT
AAAAATAAATAAAACAGCAGCAGCAAAACAAATAAAATGAAATAAAAGAAACAA
ATGAAATAAATATTGTGTTGTGCAGCATTAAAAAAAATCAAAATAAAATTAAT
GTGAGCAAAG

Sequence ID - 1148

nt: 587

TGAAAAATAAAGTTTTTATGTATATTCTACATATGTATATGTTGGTAGAAAGCAA
AAACGCTAGGTAAAAATAAATGTAATACAATTTTAGCTATGAACCAAAAAACCA
TTTGTGGTGTGGATGCAAGAAAGTCTGGATGGGTGCAGAGTTCTCCATGTTTCAC
TTCTGACATTTGAAAATACGCAGTTTGCAATTGATACGTCAAATGTTATTTTAAAG
AAAACCAATAAAATCATTAAAACCGAAAAGGCAGTTTTGCTTGTTTTACCTTAG
TTGGAGTTATCTGCAATTGCCGTATTAGTGTTTTAAGGAACTTGTAAGTAAGCTCC
TTAGTCCCCTTTAGAGCTACGAAACATGTCAATTTTACTTTTCTCCAGCTTTTGG
AATCTTATCTAAATTACCATGTAGAGTTCTGCATAGCTTCAAATTCTCTTAGCCAA
TGTGGTCTGTAAGTGCTATCGATGAATTTACCGTTAATTGCCGTAGTATACTGT
CCTGTACCGGATGTGAAGAGGAGCAACTCTGCACAGTCACTGGTTGCTCCCATG
GTAGGAANGAATGGCTTATCAATGGTCGGATT

Sequence ID - 1160

nt: 650

GGAGGATGGAGCAGTGAGCGGGTCTGGGCGGCTGCTGGCAGCGCCATGGAGACG
GTACAGCTGAGGAACCCGCCGCGCCGGCAGCTGAAAAAGTTGGATGAAGATAGT
TTAACCAACAACCAGAAGAAGTATTTGATGTCTTAGAGAACTTGGAGAAGGG
TGAGTGTAAGAACTATAGGTAGGTCAATTGGGTCCCAGTCTTTTCCCTGCCCA
GAAGAAGCAGAAGGATATGAACCTTTCAGCATTGTTCTAGGTGGGGTGGAAGGT
AAATTTACAGCTTGTGATGTCCTTCTTCGCTTACTCCAATCCCTATTATAGACAG
ATTTAGTGATTCTGCTTTTAAACACGAAGAATATCTATTGTTTTCTCTTTGTA
GGATCTGTATGATTTTATCTACTTAACAGATAGCACTAATTAGATTAAAATTCTAT
AAGAACTTTTTAATTTGCTGTTTATAATTTCTGATTGGTATGCAATAACTGTTT
AATGAAAATCAATGTAATTTAGTATTTAATAATTTGCACCTTTGTGAAATATAGTA
AATAAATTAAGCACTATCACACCTTCACAGCTACTTAGGAGATCCACAATCCTG
GGTTGGGAGCCAGTGGATTTCCTGAAACACAGATTTGTTAATG

Sequence ID - 1165

nt: 502

CTCAAGTGAATCCTGGCTTCTTGAAGCGCTTGCCTAGACGAGACACAGTGCATA
AAAACAACTTTTGGGGGACAGGTATGTTTTCTTGCAGCTGCGGTTGTAAGGTCTT
GGCAAGACAAGCAGTGTGGCCAGAATTTGAACTTCTGATGAATGTGTAATGCAA
AGGACCTTGTACATTTTTTTGTTTTCAAGGTCCTCAAAATGAGCACATGAAGAGGT
TGCTGTGAACTTTAAGTGGCCCTACTGCGCAGAAGCATTGAGATGTCACCTTGAT
GATCTGTAAGGGAACCTTGCTGATTTGGGAATGTGCTTAGGGAACACACATTCCTT
TTGACAGGGTCTGTCACTGGGTGGGTGATGAATTATACAGATGACATGTGCTTTT

TTTTCTTTTTTCAACCTCAATGGTATTCCTACAGGAAATGGATAACCATTTTAACT
GTATTTTTTGCAGCCCGTACCTTCTTGGGAATACAATTGTCTAACTTTTTATTTTTG
GTCT

Sequence ID - 1172

nt: 648

CCACAATAATAAGAGAAAAACAGGAGCAAAAGGATATACAAAACCACCAGAAA
ACAAATAACAAAGTGACAGGAGTAAGTCCTTAACTGGCAATAATAACCATGAAT
CTAAATGGATTCCATTTCCCACTTAAAAGATAAAGACATGCTGAATGGATAAAAA
GCTGTCACCCAGTTATATGCTGCCTACAACAACTCACTTCACCTGTAAACATAC
ATATGGATGGAAAGAGAAGGCATGGGAAAAGATACTCTACTCAAATGAAAACAA
AAACCAAACAAAGGTGGCTATTCTTATATGAGATAATACAGACATTAAATCAA
AACTGGAAACAAACACAAAGTCATTGTATAATGATGAATTCAATTATATCATGAT
GAATTCAATTATATCCTCCTTCCTGATCAATTCAGAAAGGAGGATATAATCTTTTT
AAATATATATACACCCAACACCAGAGCATATAAATATGTAAAGGAAGATAAAGG
GAGTCCTGTGATCAAGAATAAATATAACAATTATAAATATTTTATCTAAAGTGAT
AGATAGACTGTAATACAATAATAGGGTGGTGACATTAACACCCCTCTCACATTG
GACTGATCATCTAGAAGGGAGAAAAAGCTTTATGATTGGAAAAGCCAT

Sequence ID - 1180

nt: 622

CTTTTCCTCCCGCTGTCCCCACGGGAGGGGACTGCTCTCCCCCGCTGCATCCTTT
CTGTGAGGTACCTTACCCACCTCAGCACCTGAGAGGGTGAAATAGAATTCTAACC
TCGACATTCGGGAAGTGTTTTTGAGAAGTCTCGGTTCGGTAAGGGGAAGTCTTCCAA
GTCCGTGCAGCACTAACGTATTGGCACCTGCCTCCTCTTCGGCCACCCCCCAGAT
GAGGCAGCTGTGACTGTGTCAAGGGAAGCCACGACTCTGACCATAGTCTTCTCTC
AGCTTCCACTGCCGTCTCCACAGGAAACCCAGAAGTTCTGTGAACAAGTCCATGC
TGCCATCAAGGCATTTATTGCAGTGTACTATTTGCTTCCAAAGGATCAGGCCCTG
AGAACAATGACCTTATTTCTTACAACAGTGTCTGGGTTGCGTGCCAGCAGATGCC
TCAGATACCAAGAGATAACAAAGCTGCAGCTCTTTTGATGCTGACCAAGAATGTG
GATTTTGTGAAGGATGCACATGAAGAAATGGAGCAGGCTGTGGAAGAATGTGAC
CCTTACTCTGGCCTCTTGAATGATACTGAGGAGAACAACCTCTGACANCCACAATC
ATGAGGATGATGTGTTG

Sequence ID - 1181

nt: 155

CGCCACTTATCCAGTGAACCACTATCACGAAAAAACTCTACCTCTCTATACTAA
TCTCCCTACAAATCTCCTTAATTATAACATTCACAGCCACAGAATAATCATATTA
AA

Sequence ID - 1183

nt: 479

CGTGGCAGCCATCTCCTTCTCGGCATCATGGCCGCCCTCAGACCCCTTGTGAAGC
CCAAGATCGTCAAAAAGAGAACCAAGAAGTTCATCCGGCACCAGTCAGACCGAT
ATGTCAAAATTAAGCGTAACTGGCGGAAACCCAGAGGCATTGACAACAGGGTTC
GTAGAAGATTCAAGGGCCAGATCTTGATGCCCAACATTGGTTATGGAAGCAACA
AAAAAACAAGCACATGCTGCCCAGTGGGTTCCGGGAAGTTCCTGGTCCACAACGT
CAAGGAGCTGGAAGTGCTGCTGATGTGCAACAAATCTTACTGTGCCGAGATCGCT
CACAATGTTTCCTCCAAGAACCGCAAAGCCATCGTGGAAGAGCTGCCCAACTGG
CCATCAGAGTCACCAACCCCAATGCCAGGCTGCGCAGTGAAGAAAATGAGTAGG
CAGCTCATGTGCACGTTTTCTGTTTAAATAAATGTAAAACTG

Sequence ID - 1185

nt: 628

CTTTGATTACCTTTGAGTATTAGGTTGAAAGCTTCTCTGTGCTTGATTGAACATTG
TGATGATGTTGATTGGGTCATGTCAGATTTAGACAGTGTGTGTTAAGATAAAT
GTTTAATGGCTCTTAGCAGTGTTTCATGCCTCCCTTTTCCCCTGATACTTTAAAAA
CAGAATATACAGAAAAGGGGAGTTGGGTGAAGAATCACCATATTCTCATTACCA
GAGTAGTGTCTACCAGCTGTTTTACATTTTTCTGTTTCCTTCTGTCCTTGGAATCC
TTTTTTTAGATCCTTGTAATACTAGTAAAGATATTCCACTCTGTGTTGTAAGCATT
TTCCATTTTGCTCCATGGTCTTCATAATGCCCTGTGGTCCTTTATTAAGGGGATG
CACCATGTAGAGGTGAAAGGCTTTCCTTGACTTGCCACCATTCTGTATTTTCCT
TAGAGGAGGAGGTTTCCAACATTTCTTTTTTAGAGACAGAGTCTCGTTCTGACAC
GCAGGCAGGAGTGCAGTGGCATGATAACAGCTCACTGCAGCCTCGAACTCCTGG
GCTCAAGTTATCCTCCCACCTCAGCTTCTGAGTAGCTAGGACTGCAGGTGCCTG
CCACCACACCCAGCTAAT

Sequence ID - 1186

nt: 494

CAGCCCTCCGTCACCTCTTCACCGCACCTCGGACTGCCCAAGGCCCGCCGC
CGCCTCCAGCGCCGCGCAGCCACCGCCGCGCGCCGCTCTCCTTAGTCGCCGC
CATGACGACCGCGTCCACCTCGCAGGTGCGCCAGAACTACCACCAGGACTCAGA
GGCCGCCATCAACCGCCAGATCAACCTGGAGCTCTACGCCTCCTACGTTTACCTG
TCCATGTCTTACTACTTTGACCGCGATGATGTGGCTTTGAAGAACTTTGCCAAATA
CTTCTTCACCAATCTCATGAGGAGAGGGGAACATGCTGAGAACTGATGAAGCT
GCAGAACCAACGAGGGTGGCCGAATCTTCTTCAGGATATCAAGAAACCAGACT
GTGATGACTGGGAGAGCGGGCTGAATGCAATGGAGTGTGCATTACATTTGGAAA
AAAATGTGAATCAGTCACTACTGGAAGTGCACAACTGGCCACTGACAAAAATG
AC

Sequence ID - 1188

nt: 599

GGGAGACAAGCCCAGCCTTTCGGCGAGNATACGTCTAACCCTGTGCAACAGCCA
CTACATTACTTCAAACCTGAGATCCTTCTTTTGAGGGAGCAAGTCCTTCCCTTTCA
TTTTTCCAGTCTTCCCTCCCTGTGTATTCAATTCTCATGATTATTATTTAGTGGGGG
CGGGGTGGGAAAGATTACTTTTTCTTTATGTGTTTGACGGGAAACAAACTAGGT
AAAATCTACAGTACACCACAAGGGTCACAATACTGTTGTGCGCACATCGCGGTAG
GGCGTGGAAAGGGGCGAGGCCANAGCTACCCGCGAGAGTTCTCAGAATCATGCTGA
GAGAGCTGGAGGCACCCATGCCATCTCAACCTCTTCCCCGCCCGTTTTACAAAGG
GGGAGGCTAAAGCCCAGAGACAGCTTGATCAAAGGCACACAGCAAGTCAGGGTT
GGAGCAGTAGCTGGAGGGACCTTGTCTCCAGCTCAGGGCTCTTCTCTCCACACC
ATTCAGGTCTTTCTTTCCGAGGCCCCCTGTCTCAGGGTGAGGTGCTTGAGTCTCAA
CGGCAAGGGAACAAGTACTTCTTGATACCTGGGATACTGTGCCCAGAG

Sequence ID - 1196

nt: 412

GTCGACGCGGCCGCGGTGCTGGAGNCGATCAACTCTAGGCTCCAACCTCGTTATG
AAAAGTGGGAAGTACGTCCTGGGGTACAAGCAGACTCTGAAGATGATCAGACAA
GGCAAAGCGAAATTGGTCAATTCTCGCTAACAACCTGCCAGCTTTGAGGAAATCTG
AAATAGAGTACTATGCTATGTTGGCTAAAACTGGTGTCCATCACTACAGTGGCAA
TAATATTGAACTGGGCACAGCATGCGGAAAATACTACAGAGTGTGCACACTGGCT
ATCATTGATCCAGGTGACTCTGACATCATTAGAAGCATGCCAGAACAGACTGGTG
AAAAGTAAACCTTTTACCTACAAAATTTACCTGCAAACCTTAAACCTGCAAAA
TTTTCTTTAATAAAATTTGCTTGTTTT

Sequence ID - 1199

nt: 439

CCCATCCCCCTCGACCGCTCGCGTCGCATTTGGCCGCCTCCCTACCGCTCCAAGCCC
AGCCCTCAGCCATGGCATGCCCCCTGGATCAGGCCATTGGCCTCCTCGTGGCCAT
CTTCCACAAGTACTCCGGCAGGGAGGGTGACAAGCACACCCTGAGCAAGAAGGA
GCTGAAGGAGCTGATCCAGAAGGAGCTCACCATTGGCTCGAAGCTGCAGGATGC
TGAAATTGCAAGGCTGATGGAAGACTTGGACCGGAACAAGGACCAGGAGGTGAA
CTTCCAGGAGTATGTCACCTTCTGCGGGGCTTGGCTTTGATCTACAATGAAGCCC
TCAAGGGCTGAAAATAAATAGGGAAGATGGAGACACCCTCTGGGGGTCCTCTCT
GAGTCAAATCCAGTGGTGGGTAATTGTACAATAAATTTTTTTTGGTCAAATTTAA

Sequence ID - 1200

nt: 526

CTGGAGACGACGTGCAGAAATGGCACCTCGAAAGGGGAAGGAAAAGAAGGAAG
AACAGGTCATCAGCCTCGGACCTCAGGTGGCTGAAGGAGAGAATGTATTTGGTGT
CTGCCATATCTTTGCATCCTTCAATGACACTTTTGTCCATGTCACCTGATCTTTCTGG
CAAGGAAACCATCTGCCGTGTGACTGGTGGGATGAAGGTAAAGGCAGACCGAGA
TGAATCCTCACCATATGCTGCTATGTTGGCTGCCCAGGATGTGGCCCAGAGGTGC
AAGGAGCTGGGTATCACCGCCCTACACATCAAACCTCCGGGCCACAGGAGGAAAT
AGGACCAAGACCCCTGGACCTGGGGGCCAGTCGGCCCTCANAGCCCTTGCCCGCT
CGGGTATGAAGATCGGGCGGATTGAGGATGTCACCCCATCCCCTCTGACAGCAC
TCGCAGGAAGGGGGGTGCGCGTGGTCGCGCTCTGTGAACAAGATTCTCAAATA
TTTTCTGTTAATAAATTGCCTTCATGTAAACTG

Sequence ID - 1201

nt: 613

CTTAAAGTATGCCCTGACAGGAGNATGAAGTAAAGAAGATTTGCATGCAGCGGTTT
ATTAAAATCGATGGCAAGGTCCGAACCTGATATAACCTACCCTGCTGGATTTCATGG
ATGTCATCAGCATTGACAAGACGGGAGAGAATTTCCGTCTGATCTATGACACCAA
GGGTGCGTTTGTGTACATCGTATTACACCTGAGGAGGCCAAGTACAAGTTGTGC
AAAGTGAGAAAGATCTTTGTGGGCACAAAAGGAATCCCTCATCTGGTGACTCATG
ATGCCCCGACCATCCGCTACCCCGATCCCCTCATCAAGGTGAATGATACCATTCA
GATTGATTTAGAGACTGGCAAGATTACTGATTTTCATCAAGTTCGACACTGGTAAC
CTGTGTATGGTGACTGGAGGTGCTAACCTAGGAAGAATTGGTGTGATCACCAACA
GAGAGAGGCACCCTGGATCTTTTGACGTGGTTACGTGAAAGATGCCAATGGCAA
CAGCTTTGCCACTCGACTTTCCAACATTTTTGTTATTGGCAAGGGCAACAAACCAT
GGATTTCTCTTCCCCGAGGAAAGGGTATCCGCCTCACCATTGCTGAAGAGAGAGA
CAAAAGA

Sequence ID - 1203

nt: 692

TGCAGAGGGGTCCATACGGCGTTGTTCTGGATTCCCGTCGTAACCTAAAGGGAAA
CTTTCACAATGTCCGGAGCCCTTGATGTCTGCAAATGAAGGAGGAGGATGTCCT
TAAGTTCCTTGACAGCAGGAACCCACTTAGGTGGCACCAATCTTGACTTCAGATG
GAACAGTACATCTATAAAAGGAAAAGTGATGGCATCTATATCATAAATCTCAAG
AGGACCTGGGAGAAGCTTCTGCTGGCAGCTCGTGCAATTGTTGCCATTGAAAACC
CTGCTGATGTCAAGTGTTATATCCTCCAGGAATACTGGCCAGAGGGCTGTGCTGAA
GTTTGCTGCTGCCACTGGAGCCACTCCAATTGCTGGCCGCTTCACTCCTGGAACCT
TCACTAACCAGATCCAGGCAGCCTTCCGGGAGCCACGGCTTCTTGTTGTTACTGA
CCCCAGGGCTGACCACCAGCCTCTCACGGAGGCATCTTATGTTAACCTACCTACC
ATTGCGCTGTGTAAACACAGATTCTCCTCTGCGCTATGTGGACATTGCCATCCCATG
CAACAACAAGGGAGCTCACTCAGTGGGTTTAATGTGGTGGATGCTGGCTCGGGA

AGTTCTGCGCATGCGTGGCACCATTTCCTGGAACACCCATGGGAGGTCATGCCT
GATCTGTACTTCTACAGAGATCCTGAAGAGAT

Sequence ID - 1207

nt: 642

ACGAGAAGCCAGATACTAAAGAGAAGAANCCCGAAGCCAAGAAGGTTGATGCTG
GTGGCAAGGTGAAAAAGGGTAACCTCAAAGCTAAAAAGCCCAAGAAGGGGAAG
CCCCATTGCAGCCGCAACCCTGTCCTTGTACAGAGGAATTGGCAGGTATTCCCGAT
CTGCCATGTATTCCANAAAGGCCATGTACAAGAGGAAGTACTCAGCCGCTAAATC
CAAGGTTGAAAAGAAAAAGAAGGAGAAGGTTCTCGCAACTGTTACAAAACCAGT
TGGTGGTGACAAGAACGGCGGTACCCGGGTGGTTAAACTTCGCAAAATGCCTAG
ATATTATCCTACTGAAGATGTGCCTCGAAAGCTGTTGAGCCACGGCAAAAAACCC
TTCAGTCAGCACGTGAGAAAAGTGCAGGCCAGCATTACCCCGGGACCATTCTGA
TCATCCTCACTGGACGCCACAGGGGCAAGAGGGTGGTTTTCTGAAGCAGCTGGC
TAGTGGCTTATTACTTGTGACTGGACCTCTGGTCCTCAATCGAGTTCCTCTACGAA
GAACACACCAGAAATTTGTCATTGCCAETTCAACCAAAATCGATATCAGCAATGT
AAAAATCCCAAAACATCTTACTGATGCTTACTTCAAAAAGA

Sequence ID - 1209

nt: 620

CTCTCCTGTCAACAGCGGCCAGCCTCCCAACTACGAGAATGCTCAAGGAGGAGCA
GGAAGTGGCTATGCTGGGGGGCGCCCCACAACCCTGCTCCCGGACGTCCACCGTG
ATCCACATCCGCAGCGAGACCTCCGTGCCCGACCATGTGCTCTGGTCCCTGTTCA
ACACCCTCTTCATGAACACCTGCTGCCTGGGCTTCATAGCATTTCGCTACTCCGTG
AAGTCTAGGGACAGGAAGATGGTTGGCGACGTGACCGGGGGCCAGGCCTATGCC
TCCACCGCCAAGTGCCTGAACATCTGGGCCCTGATTTTGGGCATCTTCATGACCAT
TCTGCTCGTCATCATCCCAGTGTGGTTCGTCCAGGCCAGCGATAGATCAGGAGG
CATCATTGAGGCCAGGAGCTCTGCCCGTGACCTGTATCCACGTAATCTATCTTCC
ATTCTCGCCCTGCCCCAGAGGCCAGGAGCTCTGCCCTTGACCTGTATTCCACTT
ACTCCACCTTCCATTCTCGCCCTGTCCCCACAGCCGAGTCCTGCATCAGCCCTTT
ATCTCACACGCTTTTCTACAATGGCATTCAATAAAGTGTATATGTTTCTGGTGCT
GCTGTGACTT

Sequence ID - 1212

nt: 374

AGAGCAGCAGCCATGGCCCTACGCTACCCTATGGCCGTGGGCCTCAACAAGGGC
CACAAAGTGACCAAGAACGTGAGCAAGCCCAGGCACAGCCGACGCCGCGGGCGT
CTGACCAAAACACACCAAGTTCGTGCGGGACATGATTCGGGAGGTGTGTGGCTTTG
CCCCGTACGAGCGGCGCGCCATGGAGTTACTGAAGGTCTCCAAGGACAAACGGG
CCCTCAAATTTATCAAGAAAAGGGTGGGGACGCACATCCGCGCCAAGAGGAAGC
GGGAGGAGCTGAGCAACGTACTGGCCGCCATGAGGAAAGCTGCTGCCAAGAAAG
ACTGAGCCCCTCCCCTGCCCTCTCCCTGAAATAAAGAACAGCTTGACAG

Sequence ID - 1213

nt: 567

GAATTATTGACTTTGAATTGCATTTTCAGTACCATGAAGTCAAAGTCAGTGGTGTA
TTTGCTCATTGTTCATTCTTTCTTTTCCACCAACATTACTGCCTGCAGAGCCAGA
GGTGAGTGCAGAAATCCTGTCAATTTCGTCACTTGTGGACAACCTGCAGCTTGCCA
CAGCCTACAGTTCCACCACTGTGACCTCTGAAAACCTCCTGAACAAAAGGAAGGA
GACTTGGAATCCTGAATGGGCTTGGAGACATTAAGGGAGAAGTGCCTCCCTGGA
CCAAGGCAGAATTCAATAGAACCAGCAAGAAATTTTCTATGAATGGGAAAGCA
GGTGGCAGGGGGCAGGGGTGGAAAAGCTTTGTACAGGAATTGTGGAAAAGCTTT
TGCATTATCTCTAGTCTGAAAGTCACATTTCTCAGTTCCTTTCCACTCTCTTCTGTC
AACTTGCTGTGAGTAAATGACATCTGTACCTGTGACACGGGGCCAGGGACTATCA

CCATATGGCCCCCACACATTATCTAGTACCAGCCTGCCTGGGCCATGCCTTTTCCA
GTCACGTGACCAGCC

Sequence ID - 1214

nt: 620

CTCTCCTGTCAACAGCGGCCAGCCTCCCAACTACGAGAATGCTCAAGGAGGAGCA
GGAAGTGGCTATGCTGGGGGGCGCCCCACAACCCTGCTCCCCCGACGTCCACCGTG
ATCCACATCCGCAGCGAGACCTCCGTGCCCCGACCATGTCGTCTGGTCCCTGTTCA
ACACCCTCTTCATGAACACCTGCTGCCTGGGCTTCATAGCATTTCGCCTACTCCGTG
AAGTCTAGGGACAGGAAGATGGTTGGCGACGTGACCGGGGCCAGGCCTATGCC
TCCACCGCCAAGTGCCTGAACATCTGGGGCCCTGATTTTGGGCATCTTCATGACCAT
TCTGCTCGTCATCATCCCAGTGTTGGTCTGTCAGGCCAGCGATAGATCAGGAGG
CATCATTGAGGCCAGGAGCTCTGCCCCGTGACCTGTATCCCACGTACTCTATCTTCC
ATTCTCGCCCTGCCCCCAGAGGCCAGGAGCTCTGCCCTTGACCTGTATTCCACTT
ACTCCACCTTCCATTCTCTCGCCCTGTCCCCACAGCCGAGTCTGCATCAGCCCTTT
ATCCTCACACGCTTTTCTACAATGGCATTCAATAAAGTGTATATGTTTCTGGTGCT
GCTGTGACTT

Sequence ID - 1216

nt: 484

CAACCTTAGCCAAACCATTACCCAAATAAAGTATAGGCGATAGAAATTGAAACC
TGGCGCAATAGATATAGTACCGCAAGGGAAAGATGAAAAATTATAACCAAGCAT
AATATAGCAAGGACTAACCCCTATACCTTCTGCATAATGAATTAAGTAGAAATAA
CTTTGCAAGGAGAGGCCAAAGCTAAGACCCCCGAAACCAGACGAGCTACCTAAGA
ACAGCTAAAAGAGCACACCCGCTCTATGTAGCAAAATAGTGGGAAGATTTATAGG
TAGAGGCGACAAACCTACCGAGCCTGGTGATAGCTGGTTGTCCAAGATAGAATCT
TAGTTCAACTTTAAATTTGCCACAGAACCCCTCTAAATCCCCTTGTAATTTAACT
GTTAGTCCAAAGAGGAACAGCTCTTTGGACACTAGGAAAAAACCTTGTAGAGAG
AGTAAAAAATTTAACACCCATAGTAGGCCTAAAAGCAGCCACCAATT

Sequence ID - 1219

nt: 559

CTTGGCAGCTCCGTTATGTGCCCAGCTCTTTGCAAGGGCATACTGGGAAATGAGT
GGAGATAAAGGACCCAATCATAAGCATTTTACAGTATGGATACCCCATTTTAAAA
AGGTAAACTGAGGCACAATGCAATTTTTTTTTTTTTTTAAGGAGTTTATTTGAGCA
AACAGTGATTCATGAATCAGGCAGCACCAAACCAGAAGGAGGCTTTGCTGAANA
AGGATGAGGGACAAGCATTTATAAAGTGAATGTAGATGTAATACAAAGAAAAATA
TTTGAACCGGGTGCGGTGGCTTACACTTGTAATCCCAACACTTTGGGAGGCCAAG
GCGGGCAGATCACAAGATCAAGAGATCGAGACCATCCTGGTCAACATGCTGAAA
CCCCATCTNTACTAAAAAATACAAAAATTANCTGGGCGTGGTGGTGCGTGCTGT
AGTCCCAGCTACTTGGGCGGCTGAGGCAGGANAATTGCTTGAACCCGGGAGGTG
GAGGTTGCAGTAAGCCGAGATTGCACCATTGCACTACTCCAGCCTGGTGACAGAG
AGAGACTCCATC

Sequence ID - 1221

nt: 741

AAGCAGAANTNTCTCTAAAAACATTATCTCCTTAAAATCTTGAGGTGCATATNAG
AGCCACAGGCAATCTCTGACATATAAAATTGCAGTACAGGCCTTTCAAATTTGGC
ATTTCACTGGTACAATACAACAACCAAGATATATAATAACTGTACAGTGCCTAGA
CATTCCAGTAAGAACCATTATTTCTTTAATGTAGAATGATTAATACATATTCTAC
AAGGGGCAGTAAGGTTAGTAATTCTATAGGGTATGTCCCGACATAATTTCAAAT
TGTAACAATAACACAAACAACCTTTGTTAAGGCCATGTTTTATTGCTGATTAATGGA
CAAAAGGCAATGTAATTTATTTTCAAGTATTTTCTTGAAAGTCTGTGCTCATAAAA
ATCATGAAAAGTTGGAAAGACTGTTAAATCACTGAAACTTCAAATATATCTTACA

CAATCTTGTGGTACAAAAATACAAGTTAAATATAAACATAAAGCAATCATGGTA
ATTTTATGCAAATCTGTTTTATGTGATCATCAGTTATATATAAAAGTTTCTCAGTT
CTGTTATTTGTGAAAAGATCAATACCAGATTGAATGACTACCTATTGGCAAAGGG
CCCTAAAAAGCTTACTTTAGCACTCATCTTTTACATGGTTAAATGCATTTCTAAT
TTGAGATCACCTAAACACTGGAAAAGAAAAAAATGAAAGGGCAGTATGTCCAT
AAACCAACAAATAATTTGGCTG

Sequence ID - 1224

nt: 485

CGAAATTTCTTGTGACACAGAGGAAGGGCAAAGGTCTGAGCCCAGAGTTGACG
GAGGGAGTATTTAGGGTTCACTTCAGGGGCTCCCAAAGCGACAAGATCGTTAGG
GAGAGAGGCCAGGGTGGGGACTGGGAATTTAAGGAGAGCTGGGAACGGATCCC
TTAGGTTCAAGGAAGCTTCTGTGCAAGCTGCGAGGATGGCTTGGGCCGAAGGGTTG
CTCTGCCCCGCCGCTAGCTGTGAGCTGAGCAAAGCCCTGGGCTCACAGCACCCC
AAAAGCCTGTGGCTTCAGTCCTGCGTCTGCACCACACAATCAAAAGGATCGTTTT
GTTTTGTTTTTAAAGAAAGGTGAGATTGGCTTGGTTCTTCATGAGCACATTTGATA
TAGCTCTTTTTCTGTTTTTCTTGCTCATTTCGTTTTGGGGAAGAAATCTGTACTGT
ATTGGGATTGTAAAGAACATCTCTGCACTCAGACAGTTTACAGA

Sequence ID - 1230

nt: 741

AAGCAGAANTNTCTCTAAAAACATTATCTCCTTAAATCTTGAGGTGCATATNAG
AGCCACAGGCAATCTCTGACATATAAAATTGCAGTACAGGCCTTTCAAATTTGGC
ATTTCACTGGTACAATACAACAACCAAGATATATAATAACTGTACAGTGCCTAGA
CATTCCAGTAAGAACCATTATTTTCTTTAATGTAGAATGATTAAATACATATTCTAC
AAGGGGCAGTAAGGTTAGTAATTCTATAGGGTATGTCCCGACATAATTTTCAAAT
TGTACAATAACACAAACAACCTTTGTTAAGGCCATGTTTTATTTGCTGATTAAATGGA
CAAAAGGCAATGTAATTTATTTTCAAGTATTTTCTTGAAAGTCTGTGCTCATAAAA
ATCATGAAAAGTTGGAAAGACTGTAAATCACTGAAACTTCAAATATATCTTACA
CAATCTTGTGGTACAAAAATACAAGTTAAATATAAACATAAAGCAATCATGGTA
ATTTTATGCAAATCTGTTTTATGTGATCATCAGTTATATATAAAAGTTTCTCAGTT
CTGTTATTTGTGAAAAGATCAATACCAGATTGAATGACTACCTATTGGCAAAGGG
CCCTAAAAAGCTTACTTTAGCACTCATCTTTTACATGGTTAAATGCATTTCTAAT
TTGAGATCACCTAAACACTGGAAAAGAAAAAAATGAAAGGGCAGTATGTCCAT
AAACCAACAAATAATTTGGCTG

Sequence ID - 1231

nt: 203

TTGAGGAAGGGTCTACTGTCTTTTTAAATGGCACAAATTTTAAGAGGTTTGAGAGG
TACAGTCCCTTAACCTGCCACGGGAGAGGGGGCCCCCAAACCTTTCTTCCCCCACA
CTTCTGGTTTTCTGTGTGGAGGGGGAGCAGGGATATCTAAGCTGTGGTGTGAAAG
GGTAGGAGAGATGCTGGAGGTGGGGGTGCTGTGTTCTA

Sequence ID - 313

nt: 554

CCCGGAATCGCGGCCCGCGTCGACAACAAACCTGCATGTTCTGCACATGTATCCA
GGAACCTTAAAAAAAAAAAAAGATAGTTTGTGTGTCTTAATTGAATAATAGTAGAT
TTATAGATTAAAGATCTATGGGTTTTTAATATGGATTANAAATCTGTGGGTTTTTG
ATATGGATTANAAATCTGTGGGTTTTTAATATGGATTGGAAATCTGTGGGTTTTTA
ATATGGATTAAAAACATCTGTGGGTTTTTAATATGGATTAAACATCTGTGGGTTT
TTAATATGGATTAAACATCTGGGTTTTTAATATGGATTAAACATCTGTGGGTTTTT
AATATGGGTTAAAAATCAAAAGAAAATGAACTATTTGCTCCAGTGCAGGAAAAT

ACAGGCAATACTGGATACAATTAGATGGTCAGGAGCGATAACCCGGTTGCCATTG
TTTGAAGAAGAGAATAAGGNGCTAGCATTCCGTAGATAATTTGACAGCTA
GGAAATAGGGGGAGTCTTCTATGTAGTTAGTGAAGGCTAAATGAACATTATATG
C

Sequence ID - 361

nt: 622

CTGTNATNGAATCTGCTTGTNACTNAAATGCTAAACTCAATTCTGTAATTCAATA
GGTGCACCTNTCTGAGAAACATANNAGACAATGAGGAAAAGGATTCANCATTCC
GTGGAATTTGTACCATGATCAGTGTGAATCCCANTGGCGTAATCCAAGTAAGATG
TTCACAAAGATTTGTTTTTAATGTCTAATTAATAAAATTTTAAAGGAAGAAACATT
CTAATACTTTAATTATAAAAAAGTTAACTATTTTCAAAGGTATCAAAATACAGTTA
AACCTTTTAAAATGTATATTTCTTAATATCTTGAAATTGTAATGCCTTTTTTTTTTCC
TAAATTTTTTTTTGTGCATGAAATGAGATAGTAACAGCAGATTGGGACAACAAGGTT
ATATTCTTGCTTGAATCAGGCCATGGCTTCTTTCATCCAAATTTAGACCTCATT
TATTTACTTTGTCCCTGCCTCCCATCCCTGGATATCANGTTTGTGGATATCTACAG
TTAATAGAGTGACCAAATAGTAGGAATACTGTCTCTCTATTCTGAATAAAATACT
TTGAATCAGATTTAGAAATAATGAATAAAATACAAATCACCATTGAAATTGCTCT
AATTTTGAGAGCT

Sequence ID - 363

nt: 628

ATCACNTGAGGCAAGAGTTTGAGCCAGCCTAGCTAACATGGTGAAACCCCATCTC
TACAAAAATATAAAAAATTAGCCTGGGTGGTGATGGGCACCTGTAACCCAGCTAC
TCGGGAGGCTGAGGTAGGAGAATCACTTGAACCCGGGAGATGGAGGTTGCAGTG
AGCCAAGATCGTGCCACTGCACTCCAGCCTGTGTGACAGAACAAAGACTCTGTCTC
AAAAAAAAAATAATAATAATAATAATAAAAAAGGAATAACATAGCTAGGAATA
AATTTAATCAAAGAGGTGAAAGACTTATACACTTAAAACTACAAAAAAAAAAATC
ACTGAAGGAATTATAGACCCAAATAAAAAATAATAAAAAAGACATTCTGTGTTTTA
GGGAAAGAAGACTTAATATTGTTAAGATGTCAATACTACCCAAAGTGATCTACAG
ATTCAACATAATCCCTATCAAAATTCCAACAGCCTACTTTGTAGAAATGGAAAAG
CCAATTTTCAAATTCAGATGGAATTGCGAGGGGTTCTGAATAACAAAAACAATCT
TGGGGAAAAAAAACAAAAAACAAAGTCAAAGAACTCACACTTCTCTATTATATAA
ATTTACTACAAAGTTATAGTAATCAAA

Sequence ID - 364

nt: 528

TGAACATCCAGCCATGTCATTTCTTCCATTCCCTGCCCTGGAGTAAAGTAGATTTAC
TGAGCTGATGACTTGTGTGCATTTGTACATTGCAACCTTAGCTTACCTCTTGAAGC
ATGTAGAGCATTCAACCCACCATTCAATCACTGCCTACTCCCACCACAGCTGTT
TCGTGGTCTGTCTGCTCCCTGTGCCACCCCCACCCCATCAGGTGGGCCTTTTGCAA
GTGATGAAGTCACCTGTGGGGGAAGAGCTTTCCTTTCCTCTCCTCAACTCAGAAG
GCCTCTTCTCTTGTCTCAAGAGGGTGCTGCTGCTTTCCTCTCCTTCCCCGGCCGG
CCTCCATCCCAGTTCACCTTTTCAGAAATGGCCCTCAGTCAACTCTTCCCTTTTCT
CCTGGCTTTTATTTCTCCAGTCTCTTAAGAGTATCCTTAGCTTTAAAAACAATA
ACACAGAGGATGGGTGCAGTGGCTCATGCCTGTAATCCCAGCACTTTGGAGCCTG
GGGCGGGCGGATCACTTGAGGNCA

Sequence ID - 368

nt: 329

GAAAGATCTAAAATCGACACCCTAACATCACAATTAAGAAGACTAGAGAAGCAA
GAGCAAATTCAAAGCTAGCAGAAGGCAAGAAATAACTAAGATCAGAGCAGAG
CTGAAAGAGATAGAGACACAAAAAACCAATTCAAAAAAAACAATGAATCCAGG
AGTTTTTTTTTTAAAAAGATCAACAGAATTGACAGACTGCTAGCAAGACTAATAA

AGAAGAGAGAAGCATCAAATAGACTCAATAAAAAATGATAAAGGGGATATCACC
ACCAATCCCACAGAAATACAACTACCATCAGAGAACACTATAAACACCTCTATG
CAAAT

Sequence ID - 381

nt: 534

GACTTANATCTAAATGGACCACATTCTCTACTTAAAAAAATGCTATTAACCATGT
GATCTTCTCAGTCATGAGGTAATCTGGTGACTACCCCTTCCTCAAAGCCAGTTGGG
ATATTCTTTGAATAGAGTAAAACAGTGTTTCTAGGCTGGGAGACACCAGACATAG
TTGAGGACAGAGGTGCTAGAAAATAGGAAGTTTAAAAGCATGTGCGGTGATGCT
CAGAGGAGGTAAACCCCAACCCTCATGCTCATAGCTTCCAATCATTTTCTCTAGTTC
TTAACTCTTAAATGTGAGAAATGCTTGAAGATTCTAGTCATCTGAAGAAAGTCTC
TTTATTAAAGATTTTCATAAAAGAGACCAAAGCAGACAAACAGAAAAAGACATC
TTGGGGAAAAAACAAGGATAATGGGAAGAGAAGGAAAGTTTAAAAATTATCA
ATATCCTCAGGGGGACAAAATATTATATCCTATAAAGACAGATTTTATTTTAA
AAAAATAGAAAGCAAAACAAGCTCCTAAAAATAAAGTTTG

Sequence ID - 382

nt: 444

GTTAAGGAAGTCAGCACTTACATTAAGAAAATTGGCTACAACCCCGACACAGTA
GCATTTGTGCCAATTTCTGGTTGGAATGGTGACAACATGCTGGAGCCAAGTGCTA
ACATGCCTTGGTTCAAGGGATGGAAAGTCACCCGTAAGGATGGCAATGCCAGTG
GAACCACGCTGCTTGAGGCTCTGGACTGCATCCTACCACCAACTCGTCCAAGTGA
CAAGCCCTTGCGCCTGCCTCTCCAGGATGTCTACAAAATTGGTGGTATTGGTACT
GTTCTGTGGCCGAGTGGAGACTGGTGTCTCAAACCCGGTATGGTGGTCACCT
TTGCTCCAGTCAACGTTACAACGGAAGTAAAATCTGTGAAATGCACCATGAAGC
TTTGAGTGAAGCTTTTCCTGGGGACAATGTGGGCTTCAATGTCAAGAATGTGTCT
GTCAAG

Sequence ID - 390

nt: 523

GAATCCCTAGAAAAAGAGAATTCCCAACTTGATGAGGAAAACTTAGAACTGCGA
AGGAATGTAGAATCTTTGAAGTGTGCAAGCATGAAAATGGCTCAGCTACAGCTA
GAAAACAAAGAACTGGAAAGTGAAAAAGAGCAACTTAAGAAGGGTTTGGAGCTC
CTGAAAGCATCTTTCAAGAAAACAGAACGCTTAGAAGTTAGCTACCAGGGTTTAG
ATATAGAAAATCAAAGACTGCAAAAAACTTTAGAGAACAGCAATAAAAAAATCC
AGCAATTAGAGAGTGAAGTACAAGACTTAGAGATGGAAAATCAAACATTGCAGA
AAAACCTAGAAGAACTAAAAATATCTAGCAAAAGACTAGAACAGCTGGAAAAAG
AAAATAAATCATTAGAGCAAGAGACTTCTCAACTGGAAAAGGATAAGAAACAAT
TGGAGAAGGAAAATAAGAGACTCCGACANCAAGCAGAAATTAAAGATCCACATT
TGAAGAAAATAATGTGAAGATTGGAAATTTGGAAAA

Sequence ID - 391

nt: 566

CTTTGAAGAACTTTGCCAAATACTTTCTTACCAATCTCATGAGGAGAGGGAAACAT
GCTGAGAAACTGATGAAGCTGCAGAACCAACGAGGTGGCCGAATCTTCCTTCAG
GATATCAAGAAACCAGACTGTGATGACTGGGAGAGCGGGCTGAATGCAATGGAG
TGTGCATTACATTTGGAAAAAAATGTGAATCAGTCACTACTGGAAGTGCACAAAC
TGGCCACTGACAAAAATGACCCCCATTTGTGTGACTTCATTGAGACACATTACCT
GAATGAGCAGGTGAAAGCCATCAAAGAATTGGGTGACCACGTGACCAACTTGCG
CAAGATGGGAGCGCCCGAATCTGGCTTGGCGGAATATCTCTTTGACAAGCACACC
CTGGGAGACAGTGATAATGAAAGCTAAGCCTCGGGCTAATTTCCCATAGCCGTG
GGGTGACTTCCCTGGTCACCAAGGCAGTGCATGCATGTTGGGGTTTCCCTTACCTT

TTCTATAAGTTGTACCAAACATCCACTTAAGTTCTTTGATTGTGCCATTTCCTTCA
AATAAAGAAATTTGGTA

Sequence ID - 398

nt: 512

GGGGAGCCCCCTCTTCCCTCAGTTGTTCCCTACTCAGACTGTTGCACTCTAAACCTA
GGGAGGTTGAAGAATGAGACCCCTTAGGTTTAAACACGAATCCTGACACCACCATC
TATAGGGTCCCAACTTGGTTATTGTAGGCAACCTTCCCTCTCTCCTTGGTGAAGAA
CATCCCAAGCCAGAAAGAAGTTAACTACAGTGTTTTCTTTGCACCGATCCCCAC
CCCAATTCAATCCCGGAAGGGACTTACTTAGGAAACCCTTCTTTACTAGATATCCT
GGCCCCCTGGGCTTGTGAACACCTCCTAGCCACATCACTACAGTACAGTGAGTGA
CCCCAGCCTCCTGCCTACCCCAAGATGCCCTCCCCACCCTGACCGTGCTAACTGT
GTGTACATATATATTCTACATATATGTATATTAAAACTGCACTGCCATGTCTGCCC
TTTTTTGTGGTGTCTAGCATTAACCTATTGTCTAGGCCAAAGCGGGGGTGGGAGG
GGAATGCCACAG

Sequence ID - 411

nt: 505

TGGAGCTGAAAAATTCCTATTACCTAGGGGCATCACAACGCATTGCATTTCGCCC
GTGTTTGGGATGATGCTGGTGTAACCTACTATGCTGCCAGTCATGTAAAAGTAT
AGCACACACAATTAGTAGGTAATGCTTGCAAATAATAATGAAAGACTCTGCTACT
GGTTTATGTATTTACTATGCTATACTTTTTGTCACTTACTTTAGAGTGTACTCCTACT
TTTTTTTTTTTTTTTTTTGAGATGGAGTTTCACTCTTGTCTGTAGGCTGGAGCGAA
NTGGCGCGATCTCGGCTTACTGCAACCTCCACCTCCTGGGTTCAAGCGATTCTCCT
GCCTCANCTTCCCAGAGTAGCTGAGATTACAGGCATGCACCGCCACGCACGGGTA
ATTTTGTATTTTGGTAGAGACAGGGTTTACCATTGTTGGCCAGGCTGGTCACCAA
CTCCTGACCTCAGGTGACCCGCCTCCTCACCTCCAGAGTGTTGGGATTACAGGNG
TGAG

Sequence ID - 415

nt: 596

GTATAATTGATTCTTTTGAACCTAAAGTATAAGACTTCACGATTAGAAAAAAATT
ATCCAAAGACTAATGTAATTAAGTGAGGAAAAGGTGCTGGAGGAACTGGATAAC
CACATGGAAATGTATGAACCATGACCTCTATGTCACATACTATATATAAACTTA
ATTTGAGGTGTATCACAGAGCTAACTGTGGGGGCTAAAACGTTGAAGCCTTTGGA
TGGCCGCACAAGAGATGTCTGCATTCTAAACCTTGGGGAGGGTATGAACATTTCT
TGGTAACATGCAAAAAGCACTAACTGTAAAAGAGAACAGTTGGTTCAGTTGAATTT
CATGAAACATTGTAAACTTCTGCTAAACAACCTGACACCATTAAAGAATGTGGAAAA
AGGCTGGGCACAGTGGCTCATGCCTATAATCCCAGCATTTTGGGAGGCCGGGGCG
GGAGAATCACTTGAGGCCAGGAGTTTGAACACAGCCTGGGCAACATGGCAAGAC
CCCGACTCTACAAAAATATTTTAAAAAATTAGTTGGGTGTGGTGATGCACTCCTGT
AGTCCTAGCTGCCAGGANGCTAAGGNGGAAGGATCACTTAACCCTGG

Sequence ID - 423

nt: 387

TGTTTCTCNAGGGCGAGAGGCTGTCTTANAGCACCATTCTCTGGCCCTNGTCCCAT
GAGAAGGAACCGCACTCAGGAGCCACACTCTCCCACTNCCCTTGCCCANAAAGACT
CACAGAGGGCACGGAGCTGGCTGTGGTGAGAGGAGGTCCANCAAATTCCTGTCT
GCANAAGGGTTCTGAACACCACCGCCTGGCAGCGTGCTGGAGGAGGGATTCTCT
TTTCCTCACAGCAATTCTGACCAGAAACCTGTCAAATCAGGAATGGCTAAAATAA
GACCAGGGTATGAATGACCATCAGCCACAGTAAAACCAAGGCACAGCTCTCCTG
AGCCACCCAAGCTGCTGTGGCCCAGACTGGTGACATCACCTCAGGGCAAAAAA
AAAA

Sequence ID - 429

nt: 535

CACAGTACTCCATTTTGGGGTCCAAACTGTAATGCTCAAAATAATAAATGCTTAC
ACGAAAATTATTTATTGAGAATATTCATATAAAAAATTACCTAAAGCAAAGTAAAA
AAAGTAAAATCAAGGTGGTATATTTGAAGTGAATGGTGATTGGAAATTTTATAGCT
GTAACAAAAAGAAAGAAAACAACCTTTTTTAAAGCCTCATTCTCTTTCTTTCAAA
ATGTACCTTATTTCCACACACTCTTGGGCTGACCTTTATTTTATCAATAAGCTCAA
TATTACTTTGTTTAAAAATAAGATGCTTCAGCAAAAGTCATTCTCTCTTTAACCATA
TAATTTAAAAACTCCTCTTCACGATTGATAGCAAAATCAGAAACGTTAGGGCACC
AGTGAGTTGAAAAAACTGGTCTTAAGTTGGAAAAACTATTATTAATAATATTATC
CTATCCATCCATATCTATTGAAATTGTCAGGTCCATAATTTCAATTTAATTAATTA
TAGGAAAGAAGAAAAGATAATACCCATTTGTTCTAT

Sequence ID - 438

nt: 577

GTCGACAGGGATGACATAACTATTAGTGGCAGGTTAGTTGTTGGTCACTTTCAAC
TCTGGGTTCAAGCGATTCTCCTACCTCAGCCTCCCGAGTAGCTGGGATTACAGGC
ATGCACCGCCACACCTAATTTTCTATTCTTAGTAGAGACGGGGTTTCTCCCTGTTG
GTCAGGCTGGTCTCGAACTCCCGACCTCAGGTGATCTGCCTGCCTCAGTCTCCCA
AAGTCCTGGAACCAAGACATGAGCCACCACGCCTGGCCCCCTTTAAAAATATTTT
TGCTCATTGATGATGCACCCAGTCACCCAAGTGCTCTGATGGAGATGTATAAGGA
GATGAATGCTGTTTTCATGGCTGCTAATAACAACATTCATTCTGCAACCCCCAAATC
AAGAAGTAATTTTGACTTTCAAGTCTTATTATTTAAGAAATATAATTTTGCAAGACT
ATAGCTGCCATAGACCGTGATTCTCTGATGGATCAGACAAACTAAAATGAAAC
CTCCTGCAACGTATTCATCATTCTAGATCCCTGAGGAATCGCCACACTGACTINCA
CAATGGGTGAACTGGGTTACAGT

Sequence ID - 442

nt: 606

TCGTGCCACTGCACTCCAGCCTGGACGACAGAGTGAGACTCCATCTCAAAATAAA
TAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA
TGAAAAACCTAGTTGGTATTTTGGCTATTTAATACTATAGAAATATGGTGATCTC
ATCTTTAATAGAGTGCTTTTAAGGTCCCCAGTGATAATCTCCTAAAATCATGAAC
TTAAGAATTTATAATGTTAATATGAGGAAATGAAATCTGGATTATCTCACCACAT
ATTATATAATTCATTAGTGACAGAGCAAGAACTCCAGGTACCTGTCTATTCCAT
GTTTTTCTATCTGCCTTTAAATGTTGAGATACTACCCTTATCTCATGTGAATGGA
GAACTGCCTAAAATGCTAAAACCTGACTCAGAGGCACCCAGACATAAGTGAAGT
GTGATTAGAAAATCCTGGTCAGTTGAGTCTTAGCCAAATGTGTACCTACTGTGTCT
GCCTCTATCAAGTCAATGAAAACATGATCTGAGAACTGTAAGTCCATTTATGGAA
AGGGTTGATTTANAGATATTTTGAACCTNCAGTGATGAGCCCCTTCTCAAATAG

Sequence ID - 448

nt: 329

TACGCACACGAGAACATGCCTCTCGCAAAGGATCTCCTTCATCCCTCTCCAGAAG
AGGAGAAGAGGAAACACAAGAAGAAACGCCTGGTGCAGAGCCCCAATTCCTACT
TCATGGATGTGAAATGCCCAGGATGCTATAAAATCACCACGGTCTTTAGCCATGC
ACAAACGGTAGTTTTGTGTGTTGGCTGCTCCACTGTCCTCTGCCAGCCTACAGGA
GGAAAAGCAAGGCTTACAGAAGGATGTTCTTCAGGAGGAAGCAGCACTAAAAG
CACTCTGAGTCAAGATGAGTGGGAAACCATCTCAATAAACACATTTTGGGTAAA
A

Sequence ID - 453

nt: 747

GGATCTAAGACCAGCCTGGCAGCCACCAGATGGTGATTCTAGTCCTGGCTCAGTC
AGTAATAGGTCACTGACCCCAGAGAAATCAATTCAGCCTCCCCAGGTCTTGGAT

TTCTTTCTGTGAAAATGAAAGCATAGGTAGGAATTTCCCATGGAACAGCTAGCAG
AGGAGAAATATTA AAAAGTCAGGAGACTCATGCTATAGTTTTTCATACTTCATTACA
ACAATGTTGTTTAGGACAAGTGAGTTAACCTGTTAGCTTCCTCTATATAAAATGG
AAAGTCATTAAAAACCTACATAGCAGGGTCTTGTGAAGATCAAGTGATAATGTA
GGAAGCATGTACAAATGTCACATTCTGCCGTCACGTAATGGTCCTCACAGCTTGA
GGTAGCATTAGCATGTGTCTATGATTAGTACAAGGGTGGCAAACCTGTTGCTCTT
GGATTAAGTCTGGCTCATTGCCTGTTTTTCAAAGAAAAAAATTGTATATGTGTGTA
TATATGTTATATATAGGTACACACACATATGTGCTATATATAGCATATATACACA
CATAATATATAAACATGTACATATATAGCATTATATATATACCGTGTATAATATCT
CCAGTCCTCATGACCAGCCATGCTTGTTCATTTACATTTGCATACTCTATGATTGC
TTTCATGCAACAATGGCAGAGTTGAGTGATTGTTTTGCACAGANACTGTATGGCC
CACTAAACCTAAAAATATTAATCTCTGCC

Sequence ID - 458

nt: 682

TGCCACTGAAGATCCTGGTGTGCGCCATGGGCCGCCGCCCGCCCGTTGTTACCGG
TATTGTAAGAACAAGCCGTACCCAAAGTCTCGCTTCTGCCGAGGTGTCCCTGATG
CCAAGATTCGCATTTTTGACCTGGGGCGGAAAAAGGCAAAAGTGATGAGTTTCC
GCTTTGTGGCCACATGGTGTGAGATGAATATGAGCAGCTGTCTCTGAAGCCCTG
GAGGCTGCCCGAATTTGTGCCAATAAGTACATGGTAAAAAGTTGTGGCAAAGAT
GGCTTCCATATCCGGGTGCGGCTCCACCCCTTCCACGTCATCCGCATCAACAAGA
TGTTGTCTGTGCTGGGGCTGACAGGCTCCAAACAGGCATGCGAGGTGCCTTTGG
AAAGCCCCAGGGCACTGTGGCCAGGGTTACATTGGCCAAGTTATCATGTCCATC
CGCACCAAGCTGCAGAACAAAGGAGCATGTGATTGAGGCCCTGCGCAGGGCCAAG
TTCAAGTTTCTGGCCGCAGAAGATCCACATCTCAAAGAAGTGGGGCTTCACCAAG
TTCAATGCTGATGAATTTGAAGACATGGTGGCTGAAAAGCGGCTCATCCCANATG
GCTGTGGGGTCAAGTACATCCCAATCGTGGCCCTCTGGACAAGTGGCGGCCCTG
CACTCATGAAGGCTTTCAATGTGC

Sequence ID - 460

nt: 536

CAGAGATCAAAATAGGCCTTACACAGTGCAGCGCGAATTTAAAAGATTACCCCAT
TCAGGTGTATGGATTTTGCAGTATTAAAGATGCTGCCTGGAATAGGTCATTATCTT
CTCCAAGTACTCTGTAAAGTCAATGAGTCACATAGAGTATAAGGTTTATTATCTGC
TTTTCTTTCATTAAATAAATCTTTATTGAATTTCTACTACATTAAAAAACCAAACC
AAAAACAAAACAAAACAAAAAAACACTTCCCTGAGCCATAAAGGAGAAGGTAGTT
TTGACTGGAACCTTGAAGGATGGGTAAACTTTACGCAGATAAAGATTGAGAGAA
GACCTTCCAGGTAGAGAAAGCAGTGTGGGCACAGGCAAAGATGGAAGAACACAC
GTGGCTGTGGGAAACACAGCTAGAAGCCAGTGCAGGATAGAGAGTAGGCTATGAT
GTGCAAAGGTTANACACTGGGAGAGACAGGTCCATGAGAGTAGCTTGGACTAAC
ACAGGGAGGGTTTGAATCCCAACTGGGGAACCTANAAATCAA

Sequence ID - 464

nt: 615

CGACTTTCAACCATCAAGTGAGGAATACCTTCACATAACTGAGCCTCCCTCTTTAT
CTCCTGACACAAAATTAGAACCCTCAGAAGATGATGGTAAACCTGAGTTATTAGA-
AGAAATGGAAGCTTCTCCACAGAACTTATTGCTGTGGAAGGAACTGAGATTCTC
CAAGATTTCCAAAACAAAACCTATGGTCAAGTTTCTGGAGAAGCAATCAAGATGT
TTCCACCATTA AAAACACCTGAGGCTGGAACCTGTTATTACAACCTGCCGATGAAAT
TGAATTAGAAGGTGCTACACAGTGGCCACACTCTACTTCTGCTTCTGCCACCTATG
GGGTCGAGGCAGGTGTGGTGCCTTGGCTAAGTCCACAGACTTCTGAGAGGGCCAC
GCTTTCTTCTTCTCCAGAAATAAACCTGAAACTCAAGCAGCTTTAATCAGAGGG
CAGGATTCCACGATAGCAGCATCAGAACAGCAAGTGGCAGCGAGAATTCTTGAT

TCCAATGATCAGGCAACAGTAAACCCTGTGGAATTTAATACTGAGGGGTGCAACAC
CCCATTTTCCCTTCTGGAGACTTCTAATGAAACANATTTCTGATTGGCATTAAATG
AANAGTCA

Sequence ID - 473

nt: 694

TGGGCTTTGGGCTGGCTGCAGTCTGTCTGAGGGCGGCCGAAGTGGCTGGCTCATT
TAAGATGAGGCTTCTGCTGCTTCTCCTAGNGGCGGCGTCTGCGATGGTCCGGAGC
GAGGCCTCGGCCAATCTGGGCGGCGTGCCAGCAAGAGATTAAAGATGCAGTAC
GCCACGGGGCCGCTGCTCAAGTTCAGATTTGTGTTTCTGAGGTATAGGCGGG
TGTTTGAGGAGTACATGCGGGTTATTAGCCAGCGGTACCCAGACATCCGCATTGA
AGGAGAGAATTACCTCCCTCAACCAATATATAGACACATAGCATCTTTCCTGTCA
GTCTTCAAACCTAGTATTAATAGGCTTAATAATTGTTGGCAAGGATCCTTTTGCTTT
CTTTGGCATGCAAGCTCCTAGCATCTGGCAGTGGGGCCAAGAAAATAAGGTTTAT
GCATGTATGATGGTTTTCTTCTTGAGCAACATGATTGAGAACCAGTGTATGTCAA
CAGGTGCATTGAGATAACTTTAAATGATGTACCTGTGTGGTCTAAGCTGGAATC
TGGTCACCTTCCATCCATGCAACAACTTGTTCAAATTCTTGACAATGAAATGAAA
CTCAATGTGCATATGGGATTCAATCCCCACCATCGATCATAGCACCCTATCAG
CACTGNAAACTCTTTTGCAATTAAAGGGATCATTGC

Sequence ID - 476

nt: 476

CAGAATCTTTTCATAGGCTGAATGTTGCTCCACAATGTGTCCTTTGACTATCTCTG
GCTAATTATTATTTAATCTCTTCTCAGCTTTTCCAAGAACATAACGTAAACCAA
GATCTTAGGCCATTACAACTCTTTTGTA AAAAATTAATGTGGATGTGAAACGAGG
CAACAAATCCTGAAGTAGAAAGTTATTCCTGGCCAGGCACGGTGGCTCACGCCTG
TAATCCTGGCACTTTGGGAGGCCGAGGTGGGTGGATCATGAGGACAGGAGATCG
AGACCATCCTGGCCAACATGATGAAACCCCATCTCTACTAAAATACAAAAAATTA
GCTGGGCATGGTGACGCGTGCTGTAGTCCAGTTACTCGGGAGGCTGAGGCAGG
GGAATTGCTTGAACCTCGGAGGTGGGAGGTTGCAGTGTGCCGAGATCACGCTACT
GCACTCCAGCCTGGCAACAGAGCAAGACTCCATCT

Sequence ID - 485

nt: 551

TTTGGAACACAAAGTTCCCTTTTTAGAAGAATAGGTATTGAGCCCTTGAGCGTGG
GTAGAAAGATAGAGACAGAGTGATTTGCAAAATAATGGAGGATCATATTTATAT
ATGAATTTTCACTTATTTGAACTTTCAGATATCANCTTNA AAANCTTTGGTTTAAG
TAAAGTNTNTAATGAGACTCCTTGGATGAAAGTAACCAAAACCAGTAAAAATA
AGGTAATAAGGATGTAATAGTTTCTTATGGACACTCAACAGCTAGAATGCAGTTA
GTCTCAGAAAAGAATTAGAACAAATAACTGGAAGGCCATCAGGAGTCCAAAACC
ATCACTCTTTTATATTTATATTTTATTTTCTCTCTTCANATGAGCATTCTCTTTCT
ATGTCCATATGGTANAAGGCGGCAGCTCCATAGATTATGGCTTCAGATGTTACAG
TTCCGCTNAATGCAGGGACAGACTTGCTATCTTTCAGTCCCCTTACATATCCTGGG
GAGAGAGCAAATGATTGACTGGCTTGAGTCAGGTGCCCGTTCCCTTTCCAATCT

Sequence ID - 487 nt:224

GTTTGNNTGTGACCATCTGTACTTGTAATTTCTTTACNTTCATTGGTATGAAAAAT
ATGTTCTTAGAAGCANGAAAAAGAATTCAGNNTTGCTTTGTATACTAAATTAAAT
GCTGTAATTTTGATAAAATGAAAAATCTGCTTTATTGCAACAATTGGTTTCTTCC
TTGACGTCAGCCTCACTCTTGGACTTTGGTATTCAGCCNGNCACCCCTGGGAATTC
C

Sequence ID - 490

nt: 382

TTTTCTTAGAACTTTATTTTTCTGGCCAGGCGCAGTGGCTCACACCTGTAATCCC
AGCACTTTGGGAGGCCAAGGCAGGTCGATCACCTGAGGTCAGGAGCTCAAGACC
AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGG
GCGTGGTGGCGCATGCCTGTAATCCCANCTACTCAGGAGGCTGAGGCAGGAGAA
TTGTTTGAACCCGGGAGGCGGAGGTTGCANTGAGCCGAGATTGCGCCACTGCACT
CCAGCCTGGGCAACAGAGCGAACTCCATCTCAAAAAAAAAAAAAAAAAACAAC
CTTTATTTTTCTGATTTTAAAAGTAATAACTAGTTTGTAGAAACATTAAAAGT

Sequence ID - 491

nt: 382

TTTTCTTAGAACTTTATTTTTCTGGCCAGGCGCAGTGGCTCACACCTGTAATCCC
AGCACTTTGGGAGGCCAAGGCAGGTCGATCACCTGAGGTCAGGAGCTCAAGACC
AGCCTGGCCAACATGGTGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGG
GCGTGGTGGCGCATGCCTGTAATCCCANCTACTCAGGAGGCTGAGGCAGGAGAA
TTGTTTGAACCCGGGAGGCGGAGGTTGCANTGAGCCGAGATTGCGCCACTGCACT
CCAGCCTGGGCAACAGAGCGAACTCCATCTCAAAAAAAAAAAAAAAAAACAAC
CTTTATTTTTCTGATTTTAAAAGTAATAACTAGTTTGTAGAAACATTAAAAGT

Sequence ID - 500

nt: 390

GGAATATGGTCAGGATCTTCTCCATACTGTCTTCAAGAATGGCAAGGTGACAAAA
AGCTATTCAATTTGATGAAATAAGAAAAAATGCACAGCTGAATATTGAACTGGAA
GCAGCACATCATTAGGCTTTATGACTGGGTGTGTGTTGTGTGTATGTAATACATA
ATGTTTATTGTACANATGTGTGGGGTTTGTGTTTATGATACATTACAGCCAAATT
ATTTGTTGGTTNATGGACATACTGCCCTTTCATTTTTCTTTCCAGTGTTTAGGT
GATCTCAAATTAAGAAATGCATTTAACCATGTAAANATGANTGCTAAAGTCAGC
TTTTAGGGGCCCTTTGCCAATAGGTANTCATTCAATCTGGTATTGATCTTTTCACA
AA

Sequence ID - 503

nt: 109

ACATTTTCCGNCCTTTTGCCATACACAGTTACAGAGATCAGTCAAATCCATACC
ACCACTGAGATCTCATTTATTGCCACAGATGCACAAAATAAATAACCCAAAATC

Sequence ID - 509

nt: 575

TTTTTTTCTAAATGGNGATTACTAATATATGTGGAGACTATTAATCTCTTTTCTGTT
GCCATTAGTTCATTTTCCCCAAAAGCCAATACATGTTTCATTACAAAAATGAATTA
TAAATATAAGTTAAAAGAAAAACATAAAACCCTACAATCTTACCCACCCAGAC
AACTACTATTAATACCTTAGTATTAACATATACACATCATGTATATGTATAAATTT
ATCTTAAACAAAAATAAAATTATTCTTTACATATTGTTTTAAACCTATTTATCTG
GCCAGGTGCCGTGGCTCACGCTTGTAATCCCAGCACTTTGGGAGGCTGAGGCACG
TGGATCACCTGAGGTCAGGAATTCGAGACCAGCCCAGCCAACATGGTGAAACCC
TGTCTCTAATGGTTTAAATACCAAAAAATTAGCTGGGCATGGTGGCACATGCCTG
TAATATCAGCTAACATGGGAGGCTGAGGCAGGAGAATCACTTGAACCANGGAGG
GGGAGGTTGCAGTGAGCCGAAATCACACCACTTCACTGCAGCCTGGGCAACAAA
GCAAGACTGTCTCAAAAAGAAAAA

Sequence ID - 518

nt: 502

GATGCATGTCCAGCATAGGCAGGATTGCTCGGTGGTGAGAAGGTTAGGTCCGGCT
CAGACTGAATAAGAAGAGATAAAATTTGCCTTAAAACTTACCTGGCAGTGGCTTT
GCTGCACGGTCTGAAACCACCTGTTCCACCCCTCTTGACCGAAATTTCTTTGTGAC
ACAGAGAAGGGCAAAGGTCTGAGCCCAGAGTTGACGGAGGGAGTATTTCAAGGGT
TCACTTCAGGGGCTCCCAAAGCGACAAGATCGTTAGGGAGAGAGGGCCAGGGTG

GGGACTGGGAATTTAAGGAGAGCTGGGAACGGATCCCTTAGGTTTCAGGAAGCTT
CTGTGCAAGCTGCGAGGATGGCTTGGGCGCAAGGGTTGCTCTGCCCCGCCGCGCTA
GCTGTGAGCTGAGCAAAGCCCTGGGCTCACAGCACCCCAAAGCCTGTGGCTTCA
GTCTGCGTCTGCACCACACATTCAAAGGATCGTTTTGTTTTGTTTTTAAAGAAA
GGTGANAT

Sequence ID - 523

nt: 585

GATTTACTGTGGGAATTTGCTCATGCAATTATGGAAACCTAGAAAGTCCCATAATA
TGCCATCTTCAAGCTGGAATCCCAGGAAAGCAGGTGGTGTAAATTCTGAGATTGAA
GTCTTGAGAACCGGGGGAGTCAATGGTGTAACTCCCAATCTAGGGCTTAAGGCC
AAGGACCAGGGCTGCTGGTGTGCAGATGCAAATCCTGGAGTTCAAAGGATTGAG
AACCAGGAGCTCTGGTGTCTGAGGGCAGTAGAAGATGGATGTTCCAGCTCAAGA
AGGGAAAGTAAGAATCCGTCCTTCTCCACTTTTTTGTCTATTTCAGATGAGCCCT
CAATGGACTGAACGATGCTCACCCACACTGTGAGGGCTGGTCTTCTTTATTCAAT
CCACTGACTTAAGTGCTGATCTCTTCTGGAAACACCTTCACAGACACACCCAGAA
ATAATGTTCTACCAGCCATGGGCCTGTTACTTAGCCCAAGTCAAGTTGACACAGAA
AATTAGCTATCACAACATCTGTGTGTGTATATACATATGTATTTGCATGTGTGTGT
ATATATGNGTATATATATTCATGTGTGTGTATAT

Sequence ID - 526

nt: 516

CTTTTCATGGTCTCTTGTTTCAATTAATCATCTAAAATCCAAGCNCAGAGAATTCAAT
TTTAGATGGTCTCCAGAGCAGAATTTGATGTATTAATCTTAATTACAAATCATAGA
TAATTAATATTGNTTACAAAATCANAATACGATTAGAGGTAGGGATCCTGCACAC
ACCCTATTTTCTCTCCCCAGTGTTCTGACCGAGAGACTAATTAATAATTCAAGGAA
CTTACAGTGAATGANAAACCCATGGTTTTGCTTAATTATCAGAACAGCTAGATCTG
AGAACAGCTGTCTCCACATGGATAGACACTTATTCCACCCATTGTCAGGTAGAA
TAGCTGGCAATAATAAGTCCTTCCCATTTGGATATGTTGAAAGGTGCCTGCCATGG
CATAGTTGCCACAAGAGAGGAAGAAATGGACACAAATGTAGGCTGTTTTCAGGG
CANAGGGAAGGTGGGAGGAAACCAANTTGCTGGTTTTTCACACACCCTCTGGGGA
ACACCCATGCACCTATGANATG

Sequence ID - 562

nt: 580

ATTGCATGCAAGTTTGCTGAGCTGAAGGAAAAGATTGATCGCCGTTCTGGTAAAA
AGCTGGAAGATGGCCCTAAATTCTTGAAGTCTGGTGTATGCTGCCATTGTTGATAT
GGTTCCTGGCAAGCCCATGTGTGTTGAGAGCTTCTCAGACTATCCACCTTTGGGTC
GCTTTGCTGTTCTGTGATATGAGACAGACAGTTGCGGTGGGTGTCATCAAAGCACT
GGACAAGAAGGCTGCTGGAGCTGGCAAGGTCACCAAGTCTGCCCAGAAAGCTCA
GAAGGCTAAATGAATATTATCCCTAATACCTGCCACCCCACTCTTAATCAGTGGT
GGAAGAACGGTCTCAGAACTGTTTGTTCATTTGGCCATTTAAGTTTAGTAGTAA
AAGACTGGTTAATGATAACAATGCATCGTAAAACCTTCAGAAGGAAAGGAGAAT
GTTTTGTGGACCACTTTGGTTTTCTTTTTTTCGTGTGGCAGTTTTTAAGTTATTAGTT
TTTAAATCAGTACTTTTAAATGGAAACAACCTTGACCAAAAATTTGTCACAGAAT
TTTGAGACCCATTAAAAAAGTTAAATGAG

Sequence ID - 564

nt: 671

GGAAATAGAATTTTAAATAGTAATAACTGCTTGTTTTTTTTGTGCAAGTACTTTTAT
ACATAAGATAAAACAAAACCTTACCACCAAACATAACCAAATGCACCTCTTTCAT
AAGTGAGTTACTAAGATTTCTATACCTGGAATATCATGTATGTTTCATTTACTGGA
TGTTTACATTTTAGGAAGGAAAATAGTTTTGTTTATTTAAACAACCTGAATACTTAT
AAACTGTTGTTCTGGAAGTTATTTATTCATAAAAAATTTGTTCTTTTGTTCATGA

ATTTATAATTCCTAAATGAAGACCAGAAAGTACAAATTGCTGGGAGGAAGAATA
GGCTTTATTAATCAACTGATGTCTTGATTTTTCTAAATGGGAAGATTGCTTTATTT
TTAACTACTAATTATGGGAGCAGATTCTTAGCAAACCTTCTTTGGAAAAGTTAATGT
TATGATGTGCATTAGGCTGCCCATCGTGTATATAAATGAAGCAGATTGATTTTT
GTATTCTTACGTTTTCTCTGCTTTGTAGTTGTGGCTGTACTTAAAGAAATACAGAAT
TTCATATATTTAAAAATGTTTAAAAATGTGACCCACAGACATTGTAAATGGATTNA
AACTAACATGAAAAATATTCAACCTAAAAGAATTCTTAACTTCACAAGTGTTTT
ACTTC

Sequence ID - 575

nt: 209

CAGGATATCGAGACCATCCCAGACAGCATGGTGAAACTCCGTCTCTACTGGAATA
CAAAAAGTTAGCCGTGTGTGGTGGCACGCGCCTCTAATCCCAGCTATTCGGGAGG
CTTAGGCAGGAGAATTACTTGAACCCGGGAGGCGAAGGTTGCAGTGAGCTGAGA
TCGCACCATTGCACTCCACCCTGG-CGACAGAGCAAGACTCCGTCT

Sequence ID - 579

nt: 502

CGAATAGCCAAGTGGTCTGACAAAGATCGAGAGTAATGAGGCCCACTACTTTAGTAC
AGTCTTGAATGGCCAGATGGTGTCTGGGCATACCCCAACCAGAGATATGTAAGTCT
TTATGTTGTCAAAATTTCCAGAAACATGAATTTCCCACTAAGATTTCATTAAGGA
AACTAGAAATGAAAACAAAAACGTTCTTGTATAATATTCATTANAAAGAAATG
AAGAAGGCCGGGCATGGTGGCTCACGCCTGTAATCCAGCACTTTGAGAGGCCA
AGGTAGGCAGATCATGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATAGTGA
AATCCCGTCTCTACCAAAAAATACAAAAAAATTAGCCGGGCATGGTGGCACACAC
CTGTCATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAG
GTGGAGGTTGCAGTGAGCTGAGATTGCACCACTGTACTACAGCCTAGGTGACAGT
GCAAGACTCTG

Sequence ID - 580

nt: 316

CCTATGCCAAACTAAAGAAAGCTTGCCTGGCCTACAGGCCTAAAGGTTCAAATGN
GGATTAAAAAAACACAGTAGTCACATAAAATGTCTGCTGGCTGGCTGGAATTCCA
TCACCTACAATTTACCTGCTTTCAAAAACCTGTGTTCAACATTGAGAAAACAGAAA
ACCACTTATCTTGAGCTTAATATGGGCTTCTTTTTCTTAACTGTAGAACACTTAC
TGAAATATCAAATCAATGGTTAGGATATGTATCCTAGGCAGGCCTAAACCATTAA
CACTTGGTTTAAGCAACTTTGTATAATTNACCTCCTAAAT

Sequence ID - 583

nt: 631

CTGAGGTGGGAGGATTCCACTCTCACCCATTTCTTCTTTTCAATTTTCAGTTTCTCCA
GTTAGTAACTGAAGATGTTCTTTGAGTAATTAAGTGAGTGAGAAAATTTTAAAGT
GAGAAATCTATAAAAAGAACCATGTTAACATAAAATATTTTCACTCCTTACAAGTTG
GTATTGACTTTTCTCATTGGTAATCTGACTGATTTAATACTGCTCATTCCAATATCT
GGTGATGTAATTCTGGTTATGAATCCTTGTATTAATAACACCTCCTGGGAGGTTTT
TTTTCCCAACATTACATTCAGAATATTAGAGCTGAAAATACCTTTTTTAAGGTTA
TCAGGAGGAGGGAGCTTATGTTTAATGTGGTGGATAAACTTAACTGCTGGTTAA
TACAATTGTTATTCAGGTGAAATTCCTAAACTTTTCACGTGCAAAGTTTTGTATG
TATACAGACATTTGGGGAAAAGTTTTATCATCCCTAAAACCGGTTACTGTCCAGA
AAATGATAAGAATCCCTGGGTTCCAAATCCTTCATAAGGTATTTATTCATTTATTT
ATTCAACACATTTACTCAATGCCTCCGCTCTGCTGCAACTACACTGACATTCTGCT
TCTAATCTAACCGAAAAT

Sequence ID - 593

nt: 565

CAGGATCAAGGTGAAAAGGAGAACCCCATGCGGGAACTTCGCATCCGCAAACCTC
TGTCTCAACATCTGTGTTGGGGAGAGTGGAGACAGACTGACGCGAGCAGCCAAG
GTGTTGGAGCAGCTCACAGGGCAGACCCCTGTGTTTTCCAAAGCTAGATACTG
TCAGATCCTTTGGCATCCGGAGAAATGAAAAGATTGCTGTCCACTGCACAGTTCCG
AGGGGCCAAGGCAGAAGAAATCTTGGAGAAGGGTCTAAAGGTGCGGGAGTATGA
GTTAAGAAAAAACAACCTTCTCAGATACTGGAAACTTTGGTTTTGGGATCCAGGAA
CACATCGATCTGGGTATCAAATATGACCCAAGCATTGGTATCTACGGCCTGGACT
TCTATGTGGTGCTGGGTAGGCCAGGTTTCAGCATCGCAGACAAGAAGCGCAGGA
CAGGCTGCATTGGGGCCAAACACAGAATCAGCAAAGAGGAGGCCATGCGCTGGT
TCCAGCAGAAGTATGATGGGATCATCCTTCCTGGCAAATAAATTCCCCTTTCTATC
CAAAAGAGCAATAAAAAGT

Sequence ID - 595 nt: 98
CTTTGCTCGAATNGTCAGATAAGGATTCTGTGAANGGAGATGAGATTTCATCCA
TGCTGACTTTGANAATACATGTTCCCGAATTGGGGNCCCCAAA

Sequence ID - 598 nt: 362
GGCATGTGCCTGTAGTCCTAGTTGCTGAGGTAAGAGGATTGCTTGAGCCCAAGAG
TTCAAGGCTGCAACAAGCTTTGATTGCGCCACTGCACTCCANCCTTGGCGACAGA
CTAAAACGCTGTCTCAAAAAAAAAAACAACGACNAAAAAAAAAACAAACAG
AAAAAATTAACCTTAGGCAATGACAGTCCCTGGCAAATGCTGGGAGGGAGGCAAC
ANTGGTCAAGGAAGGTAACCCTGAANCAGGACTTGTAAGCAAATAANATTGGG
AGGCCAAGGTGGGTGGATCACNAGGTGAGGAGTTCGAGACCAACCTGGCCAACA
TAGTGAAACCCCGTCTTTCTAAAAATACAAAAAAATT

Sequence ID - 600 nt: 595
TTCAAATTCTTGNTAANAGTCTTTGTTCTGAATTTTACTTTGTCTGTTATTCCTATA
GCCTTTCCAATTTTCTTTCGCTTGGATTTTACGTGATAAGTTTTTTCCCCCATTTTA
CTTTTANCAACTCTATATTTTTTAGTTGAGGTTGGGTTTCTTGTAACAGCATATA
ATTTGGGTTTTTTAATCCAATCTGAAAATTAATGTCCTTAATTTTGTGTTTATACCA
TTTACACATAATGTACTCATATATAAGGTTTAACTGAAACCTACTATCTTGCTAGT
TGTGCTCTACTTGAATTTTTTTTAGTATTCTGTTTTAATTGACCAACATTTGACTG
TATCTCTTTGTGTAATTTTACAGGTTGCTGTAGGCATGACAATATATACACTT
AACTTTTCTCAGTACACTGAGAGTTGAAATTGTAGTACTTCGAGGAAAACATAGA
AACTTGCAATGATATCGGTTACATTTTACCACCTCCATATGTTGCAATTATTA
TGTATTAGATCTGCCTACCTCGAAAACCCATCAGTCTTTTAACTTTGCTCTCAATG
GTGATTCATATTTTTTAAAAAAACTTGAGGCAA

Sequence ID - 601 nt: 522
TCGACCGGGTTTGGAGCAGTGCCTTGTTTGTGCTGTGCAGCGGATACTCTACAGGTA
CATTTCCTTTTTTGGAACCAAAGGGAGGGATTGACAATATTGATGGTAGATCTT
TTTTCTTTAGCAAGAATTAAGGATTTTGGTGGGTGGGGGGAGGCTTCTGTGGGA
CCAAGACAATGTACTGTCAGTCAGGATTTAAGTCGAACTACCTCATCCCTTGCCC
CAGAGAACAGTTGATCGTGTTTTAAACCAAAGGTGCGGAATGGAGAGAGGGGAG
GCGGTGCATTGCAGCTTCCGATAGAGCTTTTTATTTTTGGATATCAGGAACCAATT
TTGAAGATTTCTTAAGAAAGTCATTTACATCAGGGACATGAAGAGCAAAGTAGGT
ATTTTTGGTCAGTACTTGAATTTGATAGGCTTTATGCAAACAACCTCTCCCTCTGCT
GGAGTCTGGCAAGTTTGCTTTTCACTGGACGCTAATTCAAGTGCCATACAAAAC
AAAATAAGAGTTTTACTTATAACACA

Sequence ID - 603

nt: 624

GACACACGAGCATATTTACCTCCGCTACCATAATCATCGCTATCCCCACCGGCG
TCAAAGTATTTAGCTGACTCGCCACACTCCACGGAAGCAATATGAAATGATCTGC
TGCAGTGCTCTGAGCCCTAGGATTCATCTTTCTTTTACCCTAGGTGGCCTGACTG
GCATTGTATTAGCAAACCTCATCACTAGACATCGTACTACACGACACGTAACGT
TGTAGCCCACCTTCCACTATGTCCTATCAATAGGAGCTGTATTTGCCATCATAGGAG
GCTTCATTCACTGATTTCCCTATTCTCAGGCTACACCCTAGACCAAACCTACGCC
AAAATCCATTTCACTATCATATTCATCGGCGTAAATCTAACTTTCTTCCCACAACA
CTTTCTCGGCCTATCCGGAATGCCCGACGTTACTCGGACTACCCCGATGCATAC
ACCACATGAAACATCCTATCATCTGTAGGCTCATTCACTTTCTTAACAGCAGTAAT
ATTAATAATTTTCATGATTTGAGAAGCCTTCGCTTCGAAGCGAAAAGTCCTAATA
GTAGAAGAACCCTCCATAAACCTGGAGTGACTATATGGATGCCCCCACCCTACC
ACACATTCTGAAGAA

Sequence ID - 605

nt: 338

ACCTGAGGCCTCGGTGGGGCCAGTGCGACGCTGGCTTAAGGAGCTGGAGGGGT
CCTAATACACATTTAATTCAGTTTCTCTTCCCTAAGAGGCTGCCGGAGTTGGGGCC
TCCTCCAGCAGAGACCCTCGGACCCCTGCAGGGCCTGGACTTGGGGTGAACAGG
GCTTCAGTCAGCGCAAGTATTCATTTGCATTTGGTAATTTTTCATGCCACCTATT
TATGAATATATAAATCTTTATACCAAATCTATTTTAAAACATGGAAAAGTTGCC
TTTATGGAACTTGGCAGAGCCAGAGTGTACACATTCCTAAACCATTAAACAGAT
TTCTATA

Sequence ID - 606

nt: 556

GGATAATGATACCTCTGACCTTTCTTCCTTTTGGGAAGTACTTGAGTGTGCAGCTG
CATGAGGCCTCAGCAGGAGAGAGATTTTAGGTCCAAGAAGCTATACCAGTAGGA
CAAGGCAGGAAAATACTACACTTTTCAGGATCAAGCCCCTCTGACTCTCATTGGGA
AACTGGATGTTTGCTAAGCACCTGCTTCTTAAGGATGCCGAGGGATTAAATGATA
CTCCCAGAAACCTGGAGAGATTAATGGGGCCTATGGAGAAGTGCTCTGAACTCA
GTGTTGGGACTTGAATAAAATTAACCATTTGTCATGTTTTCAGAACAATAAGCTG
TTTTATATTTTCATGTGCATGAAAGCCCTAGAATAAGTTGTGTTATTTCCAGAAAT
GAAATAGATCCACAGTTAGATGATGTGGCCATTAGGAAGTACCAAATTTATAAA
AATCACTGGAGGTCTGTCTGAGCAGTACCTAATAAAATATAGTATACTGAAAGTG
AACAGATCTTTGTCTCTTTCTTTGGCTGCTTGATACTTTATCTGTGTCTGCCGGACA
GTGC

Sequence ID - 612

nt: 576

GAGAAATATAAGATTATGTATAGATCAAATCTACCTCTATTTGGTGTCTGAAAG
AGATGAGGAGAATGGGACAACTTGGAAGCTTATTTCAAGATAACATTCCTGA
GAACTTCCCCAATCTTGCTAGAGAGGCCAACATTAATAATTCAGTAAATGCTGAAA
ACTCCAGTAAGATATTTCTTAAGAAAATTATTCCCAAGATATATACTCATCAAATT
ATCTAAGGTCAAATGAAGGAAAAAATTTTATAGGCAGCTAGAGAGAAATGTCAG
GTCACCTACAAAGAGAATGGCATAAGACAAAAAGTAGAACTCCAGCAGAACT
CTAAAAGCCAGAAGAGATTAGGGGCCAATATTTAACATTCTGAAAGAAATTCCA
ACAAGGAATTTTCATATCCAGCCAACTAAGCTTCATAATTGAAGGAGAAATAAG
ATATTTTCCAGACAAGCAAATGCTGATGAAATCCATCACCACCAGACCTGCCTTA
TAAGAGCTCCTGAGGGAAGCACTAAATATTGAAAGGGAAGAACTTTATGAACCA
TTTCAAAAACACATTTAAGTNCACAAAGCAG

Sequence ID - 613

nt: 341

CCTTATTTTACAGGTGAAAAACCACGAATCAGATAGATTTTATTTGCCCAAGTC
ACATAATATTAAGAACAGGCCAAGTGTGGTGGCTCATGTCTGTAATCTGAGCACT
TTGGGAGGCTAAGGCGGGTGGATTTCTGAGCCTAGGAGTTTGAGATCAGCCTGG
GCAACATGGCGAAACCTCATCTCTACAAAACATACAAAAATTAGTCAGTGTGGTG
GTGAGAGCCTGTAGTCTGGCTACTCGTGAGGCTGAGGTGGGAGCATCACCTGAG
CCTGGGAAGTCGAGGCTGCAGTGGCAACAGAATGGGTAACCTGGACATCAGAGT
GAGACCCGTGTCT

Sequence ID - 615

nt: 379

TAAATTTAAACATTTTAATTAGCTGGCATGATGGCATGCACCTGTAGTCCTACCT
ACTTGGGAGGCCAAGGCAGGAAGATTGCTTGAGCCCAGGAGTTTGAGCTTACTGT
GAGCTGTGATCACACCACTGCACTCCAGCCTGGGTGACAAAGGAAGACCGTATTT
CTAAAAAATAAAAAATACAAATACAACTACAACTAGCACTAGACCAACAGTGA
CTATGTACCATGAACTGAGGAATATTATTAATTCCACCATTTGCATCTGAGGTAA
CAATATGTCAATGACTTAAATAACATCATATCTCTGAGAGTAATTTCTCCTATATT
TCCATGACAAATGTTAGATAATTTTCCATTTTTCATTCAACAAAA

Sequence ID - 618

nt: 598

GATTAACTTTCATTTTAAGCTCTTCTCTACTAATTCTGTTTCGTATGTTTATTCATTT
TGCGTTGATCATATTTTGTACACCAGGCACTCTTCTCAGTTTTATATGTGTGTTAA
TTTACTCCTTTCAAGAGCCCTATGATACATGAATTTATCTCCATTTTATAGATGAG
GAAATTAAGACCTAGAGTTACTGAACCTGCCCAAGGTTATACAGCTGATGGGTAG
GGCCAGAACTTTGCCTCAGAGAATCTGAATTTCCAAAAAATAACCTAAAAGAGA
AATTTAAGTACTAATTAGTAAGCAAAGAAATGCACATTTAAGGAAGACAGTGCA
CATTTAAGGAAGACAGTAACCTTTTATCTATTAGAGAAAAACACACATTCTGTCT
TTAACACACACATAAATCTTATATTGGCAGGGATTTTCTTTATTTCAGCAATTATTT
ATTGGTTGTCTGCTTTGTGGTACACATAAATGCTGGGGATAAACACTTAATAAAA
TATACTTCCTTCTCTTGAATATCTTGCACTTTAAGTGGGAAGGTAAAGTCAACAGAG
TAGAGGTGATATATCCAAGTGATAGACTGTTTCATTGCCAGTAG

Sequence ID - 634

nt: 511

TTTTTTAATTTACCAAAAATTTGTTGACGTCCCTTGATTTGCTGATAGGGACAATA
ATTAAATATTTTCCACTTGTTTTTATAAAAACGTGAATGGTGATTTGTTTAAACAGA
TGTTGACTTAGCACCTTCTCTCTTTTTTTTTTTTTTTTTTTGAGTTGGAGTCTTGCTC
TGTCACCCAGCTGGAGTGCAGTGGCACGATTTGGGCTCACTGCAACCTCCGCCTC
CCAGGTTGGGCGCTTCTCCTGCCTCAGCCTCCCANATAGTTGGGATTACAGGTG
CATGCCGCCACNCCTAGCTAATGTTTTTGTATCTTGGTANANATGGNGTTTCACC
TTGTTGCCCATGCCGCTCTTGAACCTTGGCCTCCCAAAGTGTTAGGATTACAGG
CGTGAGCCACTGTGCCTGGCCCCAATTTANACCTTACTGGGTGCTGAGGCTGTG
AGCCATAGTAGAATGCATGTGATCCAGGGCCTTGCTGAATTCATGGGCTAATAGG
GAGCCTGAC

Sequence ID - 636

nt: 572

CTTANAAGAGTTGCTCATTACACCCACGCCCTTGCCCAAGGCTGGCCCACTCAG
AGCGAAACTTAACTTTTGTCTGGATGGGAAGAGAAGTAAGTCTACCCCGAGGTTG
CCATGTTGAAGAGTGAGAGGTCCAAGTGATTCTGTGCATTGAAACCAAGACACCC
CACCCAGAACACTTCTTCCCTCCCTCAGCCCAAACCAAAGGCTGGGGTTCTCATC
TCCAAGTGGCTGTTCTCCAACCTTCCCAAGCCGCTTGCAATCCCCAGACTGGACTA
CTGTGGCGGTTAGGTTAGATTTGAAGACGGGGCCAGGCTGGGTATGAACGGGT

GCAGCCCTCTTCTCCTCTTCCCCCCCACATCTCTCATGAGAGAGGTAGTGGCATT
CCTTCTCAGGGAGCTTCAATGGGAAAGGTCTCGAAAGCTTCAGGAGGAGCAGAA
TACCAACGCAGGGGGATGGCTGTAACGATCTCACCGTCTCCTAACCTCAGTCCCT
TTTTTGAGAGTGAATGGTGGAGGGTGGGAAAGGGACCCAAATTTGTAGATCTCTT
TGTCTGGGGGAGGGGAANGATG

Sequence ID - 638

nt: 545

TTTGAAGGCAAAGAGGGGATTAATCTGTGCTGGCATCATGTAAGGAGACTTGATAG
ATAAGAAAAAGCTTTACCTAAGTTTTGAAGAATAGGTTTTTCATAATGGAAAATT
TAAGGGAAAAATCTCCAAAAAAGTGCTACTCAAGTTTTATCCATTTGTATTTCCA
ACACAGCCTAGGACAGTACCTGCACATAGTAGGTGATTAATAAAAAATTTAGAAA
GCATTAATACTAAAGAGGAAAAATAGCAATGGCAAGAAAAACACATGTAGGGAAC
ACATGTAGCCAAAAAATAATATATAATCAGAGAAATAATAGGACTTCTGGAAAA
AAAAGATGAGATCAGATTGGTTAGGATCTTTACTAACATGACAAGAGCATGAATT
TTTTTCTGTAGATAATAAGTATGAAAGAATTTTAGCTTAAAAATTAGCATAATTT
GGATCCACATATGCAAATCAATGAATGTAATTCATAATATAAACAGAACTAAACA
CAAAAACCACGTGATTATCTCAATAGACACAGAAAAGGCCTTCAAAAAAATT

Sequence ID - 645

nt: 649

CTACAGCCTGGGCAGCGCGCTGCGCCCCAGCACCGAGCCGCAGCCTCTACGCCTCG
TCCCCGGGCGGCGTGTATGCCACGCGCTCCTCTGCCGTGCGCCTGCCGAGCAGCG
TGCCCCGGGGTGC GGCTCCTGCAGGACTCGGTGGACTTCTCGCTGGCCGACGCCAT
CAACACCGAGTTCAAGAACACCCGACCAACGAGAAGGTGGAGCTGCAGGAGCT
GAATGACCGCTTCGCCAACTACATCGACAAGGTGCGCTTCTGGAGCAGCAGAAT
AAGATCCTGCTGGCCGAGCTCGAGCAGCTCAAGGGCCAAGGCAAGTCGCGCCTG
GGGACCTCTACGAGGAGGAGATGCGGGAGCTGCGCCGGCAGGTGGACCAGCTA
ACCAACGACAAAGCCCGCGTCGAGGTGGAGCGCGACAACCTGGCCGAGGACATC
ATGCGCCTCCGGGAGAAATTGCAGGAGGAGATGCTTCAGAGAGAGGAAGCCGAA
AACACCCTGCAATCTTTCAGACAGGAAATCCAGGAGCTGCAGGCTCAGATTCAGG
AACAGCATGTCCAAATCGATGTGGATGTTTCCAAGCCTGACCTCACGGCTGCCTT
GCGTGACGTACGTANCAATATGAAAGTGTGGCTGCCAAAACCTTGACAG

Sequence ID - 646

nt: 600

GAGATGTCTCGCTCCGTGGCCTTAGCTGTGCTCGCGCTACTCTCTCTTTCTGGCCT
GGAGGCTATCCAGCGTACTCCAAAGATTCAGGTTTACTCACGTCATCCAGCAGAG
AATGGAAAGTCAAATTTCTGAATTGCTATGTGTCTGGGTTTCATCCATCCGACAT
TGAAGTTGACTTACTGAAGAATGGAGAGAGAATTGAAAAAGTGGAGCATTTCAGA
CTTGTCTTTCAGCAAGGACTGGTCTTTCTATCTCTTGTACTACACTGAATTCACCC
CCACTGAAAAAGATGAGTATGCCTGCCGTGTGAACCATGTGACTTTGTACACAGCC
CAAGATAGTTAAGTGGGATCGAGACATGTAAGCAGCATCATGGAGGTTTGAAGA
TGCCGCATTTGGATTGGATGAATTCCAAATTCTGCTTGCTTTTAAATATTGA
TATGCTTATACACTTACACTTTATGCACAAAATGTAGGGTTATAATAATGTTAACA
TGGACATGATCTTCTTTATAATTCTACTTTGAGTGCTGTCTCCATGTTTGATGTATC
TGAGCAGGGTGCTCCACAGGTAGCTCTAGGAGGGCTGGCAACTTA

Sequence ID - 651

nt: 252

CTTTGGGAGGCCGAGGCGGGCGGATCACTTGAGGTCAGGGGTTTCGAGACCAGTC
TGGCCAAACATGGTGAAACCCCAACTCTACTAAAAATACAAAAGTTAGCCAAAGTGT
GGTGGCAAGTGCTGTAAATCCCAGCTACTCGGGAGGCTGAGACAGGAGAATCAC

TTTGAACCTGGGAGGCGGAGGTTGCAGTGAGCCAAGATCGTGCCACTGCACTTCA
GCCTGGGCAACAGAGCAAGATTCCGTCATCTC

Sequence ID - 663 nt: 627

GCCTCCCGGGTTCAGGGATTCTCCTGCCTCAGCCTCCTGAGTGGCTGCATTGCAG
GCACCTGCCACCACGCCCTTGCAAATTTTTGTGTTTTAGTGAGATGGGGTTTTGC
CATGTTGGCCAGGCTGGTCTCGGACTCCTGACCTCAGGTGATCCGCCCGCCTCAG
CCTCCAGAGGGCTGGGATTACAGGCGTGAGCCACTGTGCCTGGCCCCAAGTTTT
GCATCTTTTAATGCCCTCTGAACAAATACATAGAGAAAACCTCTCAGAACAATTAA
AACCTGCAGAGCAACAGTGTCTCCATGTCTTAGGTTTCAAGTTTGCCTCTAAAAT
TCTAATCCATATTTTTCTACTTCTCAGATAATTTATGTGTGTGTACTCTTCCTAGAC
GTACAAGAGACTTTTTAATGCTAAATATTTGTGAGTCTTAACAAAAACTCAATTT
CACATTACTCATATTGTTTTTTGTTTTAATTGAATGTGAATTAAATTTTTATTAGTTA
TTTGATTGGAATGTTATGTATGCCATTAACACTATTAGGGGAATCTCTAGCATT
CTGTATTTTTAAAGAATTTGATTCTTTTGTANATTCTGCCTGTGTGGCATTTTAAAC
ATGTGTGACAT

Sequence ID - 666 nt: 252

ATAATTCAGAACTTCTTCATATGCTCGAGTCTCCAGAGTCACTCCGTTCTAAGGTT
GATGAAGCTGTAGCTGTACTACAAGCCACCAAGCTAAAGAGGCTGCCAGAAA
GCAGTTAACAGTGCCACCGGTGTTCCAAGTGTAAAAATTGATCAGGGACCATGA
AAAGAACTTGTGCTTACCGAAGAAAAATATCTAAACATCGAAAAACTTAAAT
ATTATGGAAAAAAACATTGCAAAATATAAAAT

Sequence ID - 687 nt: 268

TTTATGTGTTTTTGCTTGGGGGGCGCTGGGCCTAGCCCAGAGTAGTGCTTGCTCCC
CCTGCCTTGTCCCACCAGGGAGGCAGCAGACTCAGGCCCTCCATGGTCTCTTTG
TCATTTTGTGACATGCATTCTCTCTTTTGTGTCATCTTGTGGGGGGAGGGGATTAA
CCAAAGGCCACCCTGACTTTGTTTTTGTGGACACACAATAAAAGCCCCGTTTATTT
GTAAAAA

Sequence ID - 701 nt: 579

CTTTGGAGCTTCTGTCTGTGCTGTGGACCTCAATGCAGATGGCTTCTCAGATCTGC
TCGTGGGAGCACCCATGCAGAGCACCATCAGAGAGGAAGGAAGAGTGTGTTGTG
ACATCAACTCTGGCTCGGGAGCAGTAATGAATGCAATGGAAACAAACCTCGTTG
GAAGTGACAAATATGCTGCAAGATTGGGGAATCTATAGTTAATCTTGGCGACAT
TGACAATGATGGCTTTGAAGGTAATTAAAATTATCAAATTGGTGCTTGATTCTGC
TTTTAAAATGGTTTATGGAAGAAAATATGATTAAAGTTTTGTATTGTTTCCTTCC
TATAGAAGATGGAGCCAGAATGGCATGCTAAGTTTTTTCTTTTCTTTAGTGTTATA
TATGACTTCTCCTCAATTGTCACCCATTGATCTTTACCACTGTTAATAATGGATGA
TATTCAAATAACCTTATTTTCAAGTATTCTAAGGCACCATTGATTAGAACTGCATT
ATTATTTATGTGTCCCTAAAAGCTACCTATTAAGCTGTTACACCCACCATTTTCT
GTTAAGAAAATCCTGATTTCAGAA

Sequence ID - 706 nt: 496

CAACCCTCTCTCCTCAGCGCTTCTTCTTTCTTGGTTTGATCCTGACTGCTGTCATGG
CGTGCCCTCTGGAGAAGGCCCTGGATGTGATGGTGTCCACCTTCCACAAGTACTC
GGGCAAAGAGGGTGACAAGTTCAAGCTCAACAAGTCAGAACTAAAGGAGCTGCT
GACCCGGGAGCTGCCAGCTTCTTGGGGAAAAGGACAGATGAAGCTGCTTTCCA
GAAGCTGATGAGCAACTTGGACAGCAACAGGGACAACGAGGTGGACTTCCAAGA
GTACTGTGTCTTCTGTCTGCATCGCCATGATGTGTAACGAATTCTTTGAAGGCT

TCCCAGATAAGCAGCCCAGGAAGAAATGAAAACCTCCTCTGATGTGGTTGGGGGG
TCTGCCAGCTGGGGCCCTCCCTGTCGCCAGTGGGCACCTTTTTTTTTTCCACCCTGG
CTCCTTCAACACGTGCTTGATGCTGAGCAAAGTTCAATAAAGATTTTGGGAAGTT
T

Sequence ID - 707 nt: 397

CGGATGTGGTGGCAGGCGCCTCTAGTCCCAGCTACTCGGCAGGCTGAGGTAGGA
GAATGGCTTGAACCCAGGAGGTGGAGCTGACAGTGAGCCGAGATCGCGCCACTG
CACTCCAGCCTGGGCGGCAGAGCGAGACTCCATCTCAAAAAAAAAAAAAAAAAA
AATAGACTTTGAGACCAGCCTGACCAACATAGTGAAACCCGTCCTACTAAAAAT
ACAAAAATTACCCGGGCGTGGTGACGGGCGCCTGTAATCCCAGCTACTTGGGAG
GCTGAGACAGGAGAATCACTTGAACCAGGGAGGCGGAGGTTGTAGTGAAGTAA
ATCGTGCCCTGCACTCCAGCCTGGGTAAACAAGAGCGAAACTCCGTCTCAAAAA
AAATAAATAAATAAAAT

Sequence ID - 708 nt: 293

CCAGCTTTTTATGGTGTTTAATCTAATACTTAAGCTGCAGTCCCAAAATTAGGG
GTCCTTCAGTCTTGGAGACTATAAGGGAGCCTCTGCACCCAGGGAAAATGTTACC
CTTTACAGGGGGGAAGGGTAAACCAGTAGGGAATACAGTACAATCCCAACCCTA
CTGGGAGGGGCGGGAGGGAGGTGTTGCCGTCCTGTATTAAGTCGATGTTGGGA
AACGTTTTAATCTGAGCCTTTGTGGGTGGAAATATGTCTCCAGTTACAACCTCC
GCAGTGGATGTGAAGAAG

Sequence ID - 711 nt: 498

GTGGTACATATACACAAAGGAAAACCTATGTAGCCATTAAAAGAAAAGGAACTCC
TATCATTTGTAACAACATAAATAAATCTGGAGGAGATTAGGCTAAGGTGAAATAA
GCCAGGCACAAAAAGACAACCTACCATATGATCTTACTTATACGTGTGTGGAATCT
AAAAAGGTGGAATTTACAGAAGCAGAGAGTAGAATGGTGATTACCAGAGGCTGG
GGAGTGAGGGCAGGAGGTTGGAGAAATGTTGGTCAAAGGATACAAAGTTTCAGT
TATACAGGATGAATAAGTTCAAGAGATCTATTGTACAACGTGGTGGCTATAGTTG
ATAACAATGTATTGTGTTCTTGAAAAATGCTGAGAGAGTAGATTTTAAGTGTTCT
CACCACAAACATAAGTATGTGAGGTAATGCATGTGTTAATTANCTTAATTTAGA
CATTTCATAATGTATTATACATATTTCAAACACGTTGTACATGAGAAAGATAC
ACAATT

Sequence ID - 726 nt: 260

CGGGGTCTGTACCGGGCTGGCCTATCACCTCTTATGCACACCTCCCACCC
CCTGTATTCCCACCCCTGGACTGGTGGCCCCCTGCCTTGGGGAAGGTCTCCCCATGT
CCCTGCACCAGGAGACAGACAGAGAAGGCAGCAGGCGGCCTTTGTTGCTCAGCA
AGGGGCTCTGCCCTCCCTCCTTCTTGCTTCTCATAGCCCCGGTGTGCGGTGC
ATACACCCCCACCTCCTGCAATAAAATAGTAGCATCGG

Sequence ID - 865 nt: 122

CCANAATCCACTCTCCAGTCTCCCTCCCCTGACTCCCTCTGCTGTCCTCCCCTCTC
ACGAGAATAAAGTGTCAAGCAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAA

Sequence ID - 869 nt: 667

TTGTGTTTTTAGGACTCCTTATCTAAATTAAGGCAGAGAAGTTACAGTATTTATAT
CTGCATTAAATCTCAATTCCAGAAAAACCTTTTAAAAAATTATTTAATCCTCTGGA

nt: 642

GCAAGTCTTCAGTAIGTACATTTATCCCTAGAAGAAGAAAAATTAGTTGTGCAT
GAAAAAGAAACATTAACTGCAAAGCTAAATGCTCACACTCTAAATCAGTGCTCTC
CAAAGTACAGCAGGCGGGAAAAAGAAAATGGTAGATTTTTTTCTTCCAATTACTTI
AACTTATTCTTTTTAATGGACACTTCATACATAAAATATATTACAAATATATTAATA
TATACATAATGTATAAGCATACATATTGAATGTGCAGTCAAAAAATGTACTAATG
GAATGCTCTACCAAAACAAGTTCACGTTTCATCTGTAAAAATGGGAATAATATTTTT
AAAAGGCATACAGTCTGAACATTTTTAGATTATTCATAAAATCTATTTCAGAAAGT
TAACTAAAAAATTTAACGTATGCCTATAACAAATTTTGTACTTAATGTAATTGNT
TTTCATCCTGAGATCTAATATCCTCGTTTTTAAAGTAGAGCCACTTGTTTGCTACAG
TTTAGTCAAAACGTTAACATTAGATGGGTAAGTAATATGAAATCTTTCTACTAC
TCCAAAATAGAAAACAGAACATTAAAAAGATAAAAAATTCAAACATACTTACCAG
TAGATTTTTCAACTGNGCAAAAGCTCATTGCATGGG

nt: 115

AAACTTTTGTGGCAACAGTGCCTAATTTGGATAATGTTTGTTCCTAATAATTA
GAGCCAAATTGTAAA
AAAAAAAA

DL: 634

GCCAGGCTTTGTGAATTACAGGACATTTGAGACAATCGTGAAACAGCAAATCAA
GGCACTGGAAGAGCCGGCTGTGGATATGCTACACACCGTGACGGATATGGTCCG
GCTTGCTTTACAGATGTTTCGATAAAAAATTTTGAAGAGTTTTTTAACCTCCACA
GAACCGCCAAGTCCAAAATTGAAGACATTAGAGCAGAAACAAGAGAGAGAAGGT
GAGAAGCTGATCCGCCTCCACTTCCAGATGGAACAGATTGTCTACTGCCAGGACC
AGGTATACAGGGGTGCATTGCAGAAGGTCAGAGAGAAGGAGCTGGAAGAAGAA
AAGAAGAAGAAATCCTGGGATTTTGGGGCTTTCCAATCCAGCTCGGCAACAGACT
CTTCCATGGAGGAGATCTTTCAGCACCTGATGGCCTATCACCAGGAGGCCAGCAA
GCGCATCTCCAGCCACATCCCCTTGATCATCCAGTTCTTCATGCTCCAGACGTACG
GCCAGCAGCTTCAAAGGCCATGCTGCAGCTCCTGCAGGGACAAGGACACCTAC
AGCTGGCTCCTGAAGGAGCGGAGCGACACCAGCGACAAGCGGAAGTTNCTGAAG
GAGCGGCTTGCACGGCTGACGCAGGCTCGGCGCCG

nt: 626

GGCAGAGGTTGCAGTGAAGTGAAGATCATGCCATTGCAATCCAGCCTGGGCAACA
AGAGTGAGACTCCATCTCAAAAAAAAAAAAAAAAAAGACAAGAGTTTCCACTCTA
AACACTTNTATTCAACATAGTCCTGAAAGTCGTAGCCACAGCAATTTAACAAGAT
AAAGCAATAAAATGTATTCAAATAGAAAAAGAGGAAGTCAAATTATCTTCACTG

GTGATATAATTCTCTACCTGGGAAACTTCACCGAAAAAGATTTACCCAAAAGATT
TCTAAGCCTAAATAATGACTTCAGCAAAGTCTCACCATACAAAATCAACATACAC
AAATGAGTAGCATTCTGTGCACCAATAATATTCAAGCTGAGAAAAAAGAACA
TGGTTCTATTTACAATAGCTACAAACAAAAAATATGTACCTAGTAATACATTAA
ATCAAGGNGGTAAAATATCTNTACAACAAGAACTACAAAAGCTGCTGAAAAA
TAGAGACACGCAAATAAGTAAAAAGGCACTCCATGCTCATGAATTTAAAGAATC
AATATAATTAAAATGTCCGNGCTGCCTAAAGCAACTTACAGATTAAAGGCTATTT
CTCTCAAACCTATAAATGCACCTTTTAA

Sequence ID - 893

nt: 585

GTCATTGCTGGGTGGCGCCAGCCCTCAGACTTGCCTCTTTGCAGTAGGAAGAAGG
CCTCCCCACATACCTTCCACACTCATCACCTTAAGCCAGACTCGGTGTCCAGTGA
ATATGACCATCTCTTGCCCATTTTCTAATGAGTGTTTTCAATTAATGAGTTATAAGA
ATGTGGTGGGTAAATCTATGGGCTTTGAACTAGTGAATCAACTTGGTTTCAGAAAT
CTGGCACTGCTACTTACTAGTGAATTTAAGCAAGTTATTTACCTTTTCAAGAGTGTC
AGTTCCTCATGCATACAAGGAAGATAAAAAATAATGTNTACNAAAGTATTGGA
GTAATTAATACATGGAGAACTACATGTAAAGCGTTTAGCATGATGTCTGACATAT
TAAGCATCCAATATTAGTNGCTTGCAGAATTATTAGTAAAAGAGATTGCTTCTGA
AAGCCATTCCAATTCTTAAATTTTATAATGCCACATTTGAGGTCACCTGAAGTCGT
GTATAACATGTGTACATTTTTCGATTATTTTTCATTCCCANATTAAAGGCAT
AGAGATATCCTAGCNANGGACTCCAAGTGTG

Sequence ID - 895

nt: 560

GTAATTGCAGCCTGGGCAACGGAGTGAGAGACTGTCTCAGGAAAAAAAAAAGAA
AAAAAACTACTGAGGTAGTTGAATATATCCTCCATTCCCCATTTGTGGATTAGTTA
GTAAATGGGGCATCTTAGGGTTTAAATATGTCCAGGGTCACTGAGGATCAGATCC
TAGGGTTCTTTGACTCAAGGCTTTTGTCTCAGCAAAACGTCACCTTCCAGCAGG
AAGGCTTTCTCAGGCAAGTAGCAGGGTGGCTACTATGTATCGCTTCTTTATTTTTT
CTTTTTTAAAATAATGCAGGCACCGTGCGCATAATTTAAAAAATCAGTGCTAAAA
CCCTTAAAAAAGCTGTTCTCATCTCCTGTCTTTCTTTTTTTTTCTTTTTAT
TTTTTTCTTTTATTATTATTATACTTTAAGTTTTAGGGTACATGTGCACAACGTGCA
GGTTTGTACATATGTATACATGTGCCATGTNGGTGAGCTGCACCCATTAACTCGT
CATTTAGCATTAGGTATATCTCCTAATGCTATCCCTCCCCCCTCCCCCTTTTTTTT
TT

Sequence ID - 905

nt: 655

CTCAGCTCTTGCCCTGGTCACCTTGTGGCTTTTACCATCCTCATCCCCTGTGCCACC
CACATCCTGCCACTTCTGCATGGAGTTGGGGTGGGGCCATTGGAGAAAAGAGGTT
AAACAAGCAGTAATTTACTTGAGTACAGTCTTTGAGCCAATGAAATGCCAGTCAT
CATTTCCCAGGGGTACTTGTATCTTGTCAACAACCCGCTGATAATGCTCCTTCAA
TGTGAATAGCAAAAGTAGGGAGAGACGCTGAATGAAGAAGATGCCTACCCCTCA
GGAAGACTGCTGTCCGCCTCCAGGCCTGCATGCACACACCCATGCCACCTGCAC
CCCCAGCACCACGCCCACTCACTCGCACACACCCACATGCCAGTGTTTTGGGG
TTGGCAGCCTGGACACTGCTGAGGCAAACACAAGTCATCAAGCATAATTCTCATT
CTCTCCTTCTGTCTCTGTTTTAGTTACAGGAATTTGCTCAGTTTAGAGGATTTAAT
AAGTCCGTGGAAAATTTGTTTTCTGTCTCTTGCTACCCACGTGAAAAGTAAGTGCA
TGCTTCATGATGTGTTTTCCCACTACCTTCCAGGCCAGCCGAGCCCACTGGCCANG
GCCTGGCCCGGTGACCTCGGTTGACACTGTCTCANGCCACTCACTT

Sequence ID - 907

nt: 582

CTTCCATTGGGGGTAAAGATCAAACCTTTAGGCGAGCCAGGTCTGTATCTCCATTC
CTGTCTCTGACTGCTTCCCTGTAGGGATTGTCTGCAAGCGCACACCTGCATTTTCT
TGTCACAAAGTCTATGCTCTAACTCTGTACCTGCATGGCTGCAAATTAGCTTCCT
TCTTCTGCCCCTCTTCTCTCTAGCTTGGATTTTGAATTTGAATGGCAGGCATGGGA
TGTCGGTGTGTGTGTACTGCTGATGTGTACAGCCGCTTGTAGCGCTCTCATTGTC
TTCAAATGTAAGTCAATTTTGGCTGGGTGCGGTGGCTCATGCGTATAATCCACGCT
TTGGGAGGCTGAGGTGAGCTGATCATTGAGGTTAGGAGTTCGAGACCAGCCTGG
CCAACATGGCAAAACTCCATCTCTACCAAAAATAACAAAATTAGCTGGGTATGGT
AGTGCACGCCTGTAATCCCAGCTACTTGAATGCTGAAGCAGGAGAATTGCCTGA
ACCCANGAGGCGGAGGTTGCGGTGAGCCAAGATCACGCCACTGCACTCCAACCT
GGGTGACAGAGCAAGGCTGTGTCTCAA

Sequence ID - 911

nt: 595

GAGGGTGTAGAAGAGAAGAAGAAGGAGGTTTCCTGCTGTGCCANAAACCTTAAG
AAAAAGCGAAGGAATTTTCGAGAGCTGAAGATCAAGCGCCTGAGAAAGAAGTTT
GCCAAAAGATGCTTCGAAAGGCAAGGAGGAAGCTTATCTATGAAAAANCAAAG
CACTATCACAAGGAATATAGGCAGATGTACAAANCTGAAATTCGAATGGCGAGG
ATGGCAAGAAAAGCTGGCAACTTCTATGTACCTGCAGAACCCAAATTGGCGTTTG
TCATCAGAATCAGAGGTATCAATGGAGTGAGCCCAAAGGTTGAAAGGTGTTGC
AGCTTCTTCGCCTTCGTCAAATCTTCAATGGAACCTTTGTGAAGCTCAACAAGGCT
TCGATTAACATGCTGAGGATTGTAGAGCCATATATTGCATGGGGGTACCCCAATC
TGAAGTCAGTAAATGAACTAATCTACAAGCGTGTTATGGCAAAATCAATAAGA
AGCGAATTGCTTTGACAGATAACGCTTTGATTGCTCGATCTCTTGGTAAATACNG
CATCATCTGCATGGAGGATTTGATTTCATGAGATCTATACTGTTGGAAAAC

Sequence ID - 912

nt: 651

CATTTCCAGAGTTTATGTGAATTGAATTGAACCTATGGTTTTATGTTACTGTCAGTA
GAATGAAGTACGAATATTTGAAAAATACACCTTCAACTTCAAAGTGATTCTTGAC
AAAAATTATAAGGAATCATTTTGGACACATTTTCTGGTAGAGCCTTGTA AAAAATT
AAAACCAAGTGTTGTTTCAAGAAGAAGTGTAAATACATAATCAGGAATTTGAGTA
GGGAGATTATTTGTTATTTAAAATTAAGTGGCTGTGTAGTTTTAACTTTAGTAT
TGCAGGTAGAGTAAGCTTACATGATAACAAAAATCTTGGTCTTAGTGACTTAATG
ATTCTGATATTTATTGATTGATTGGTTATCATTCCAAATATTTTAAAAGATAATAG
CTGGCTGGGTGCGGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAGGAC
GGGCGGATCACGAGGTCAGGAGATCAAGACCATCCTGGCTAACACGGTGAAACC
CCGTCTCTACTAAAAATCAAAAAATTAGCCGGGTGTAGTGGCGGGCACCTGTAGT
CCCAGCTACTCAGGAGGCTGAGGCAGGAGAATGGCATGAACCTGGGAGGCGGAG
CTTGCAGTGAGCTGAAATCGTGCCACTGCCTCCACCTGGCGACAA

Sequence ID - 915

nt: 230

TTTGAGACCAGCCTAGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAA
ATTAGCCGGGCGTGGCGGCACATGCCTATAATCCCACTTACTTGGGAGGCTGANG
TAGGAGAATCGCTTGAACCCANANAGGCAGAGTTTGCAGTGAGCCGAGATTGTG
CCATTGCACTCCAGCCTGGGCGACAGAGCGAGACTCCATCTAAAAANAAAATAAA
TGAATAAAATAA

Sequence ID - 61 nt: 362

CTTATTGAAAATTTTACTAATTTCTTACTTTTTAGGTTTTAGGAGAATACTTTTGGG
TAATTGACTAGCCTCACATTATATTGATAGAGGTTCTTGAAAACTTTAATGCCAAT
TCATGTATCTTATGACTAAAATAGATAATCCATTTAGAAATTTAAGTCATTCTTGC
GTGCTTGATATGTGTGTCAGCACTATCCAAGTTGCTAGGGGATACAATGGTGAAGTG
AAAATATCAGCTAGGTGCCGGTGGCTCACACCTGTTATCCCAACAGTTTGGGAGG
CCAGGGTGGGAGGATCACTCAAGCACANGCGTTTCACACCAGCCTGGACAACAT
ACAAGACCCCATCTTTACCAAAAAGTTAAG

Sequence ID - 93 nt: 405

GGATCCTGTGGCCACAGAGCTGCCCCAGCAGACGCTCCGCCCCACCCGGTGATG
GAGCCCCGGGGGGACAATCGTGCCTGGGGAGGAGCAGGGTACAGCCCATTCCCC
CAGCCCTGGCTGACCTGGCCTAGCAGTTTGGCCCTGCTGGCCTTAGCAGGGAGAC
AGGGGAGCAAAGAACGCCAAGCCGGAGGCCCGAGGCCAGCCGGCCTCTCGAGA
GCCAGAGCAGCAGTTGAATGTAATGCTGGGGACAGGCATGCTGCCGCCAGTAGG
GCGGGGACCCGGACAGCCAGGTGACTACCAGTCCTGGGGACACACTCACCATAA
ACACATCCCCAGGCAGGACAGATCGGGGAAGGGGTGTGTACCAGGCTATGATTT
CTCTTGCAATTAATAATGTATTATTATT

Sequence ID - 892 nt: 559

TCTTTCGGAAGCGCGCCTTGTGTTGGTACCCGGGAATTTCGCGGCCGCGTCGACGC
GGTCGTAAGGGCTGAGGATTTTTGGTCCGCACGCTCCTGCTCCTGACTCACCGCT
GTTGCTCTCGCCGAGGAACAAGTCGGTCAGGAAGCCCGCGCGCAACAGCCATG
GCTTTTAAGGATACCGGAAAAACACCCGTGGAGCCGGAGGTGGCAATTCACCGA
ATTCGAATCACCTTAACAAGCCGCAACGTAATAATCCTTGGAAAAGGTGTGTGCTG
ACTTGATAAGAGGCGCAAAAAGAAAAGAATCTCAAAGTGAAAGGACCAGTTCGAA
TGCCTACCAAGACTTTGAGAATCACTACAAGAAAACTCCTTGTGGTGAAGGTTT
TAAGACGTGGGATCGTTTCCAGATGAGAATTCACAAGCGACTCATTGACTTGAC
AGTCCTTCTGAGATTGTTAAGCAGATTACTTCCATCAGTATTGAGCCAGGAGTTG
AGGTGGAAGTCACCATTGCAGATGCTTAAGTCAACTATTTTAATAAATTGATGAC
CAGTTGTTT

Sequence ID - 77 nt: 464

GCGGCTGCTGTTGGTTGGGGGCCGTCCCGCTCCTAAGGCAGGAAGATGGTGGCCG
CAAAGAAGACGAAAAAGTCGCTGGAGTCGATCAACTCTAGGCTCCAACCTCGTTAT
GAAAAGTGGGAAGTACGTCCTGGGGTACAAGCAGACTCTGAAGATGATCAGACA
AGGCAAAGCGAAATTGGTCATTCTCGCTAACAAGTGGCAGCTTTGAGGAAATCT
GAAATAGAGTACTATGCTATGTTGGCTAAAAGTGGTGTCCATCACTACAGTGGCA
ATAATATTGAACTGGGCACAGCATGCGGAAAAATACTACAGAGTGTGCACACTGG
CTATCATTGATCCAGGTGACTCTGACATCATTAGAAGCATGCCAGAACAGACTGG
TGAAAAGTAAACCTTTTCACCTACAAAATTTACCTGCAAACCTTAAACCTGCAA
AATTTTCCTTTAATAAAATTGCTTG

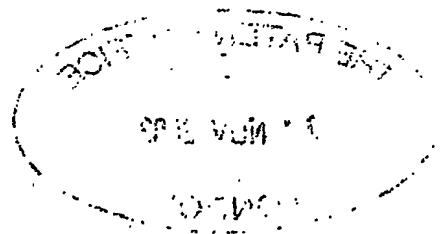
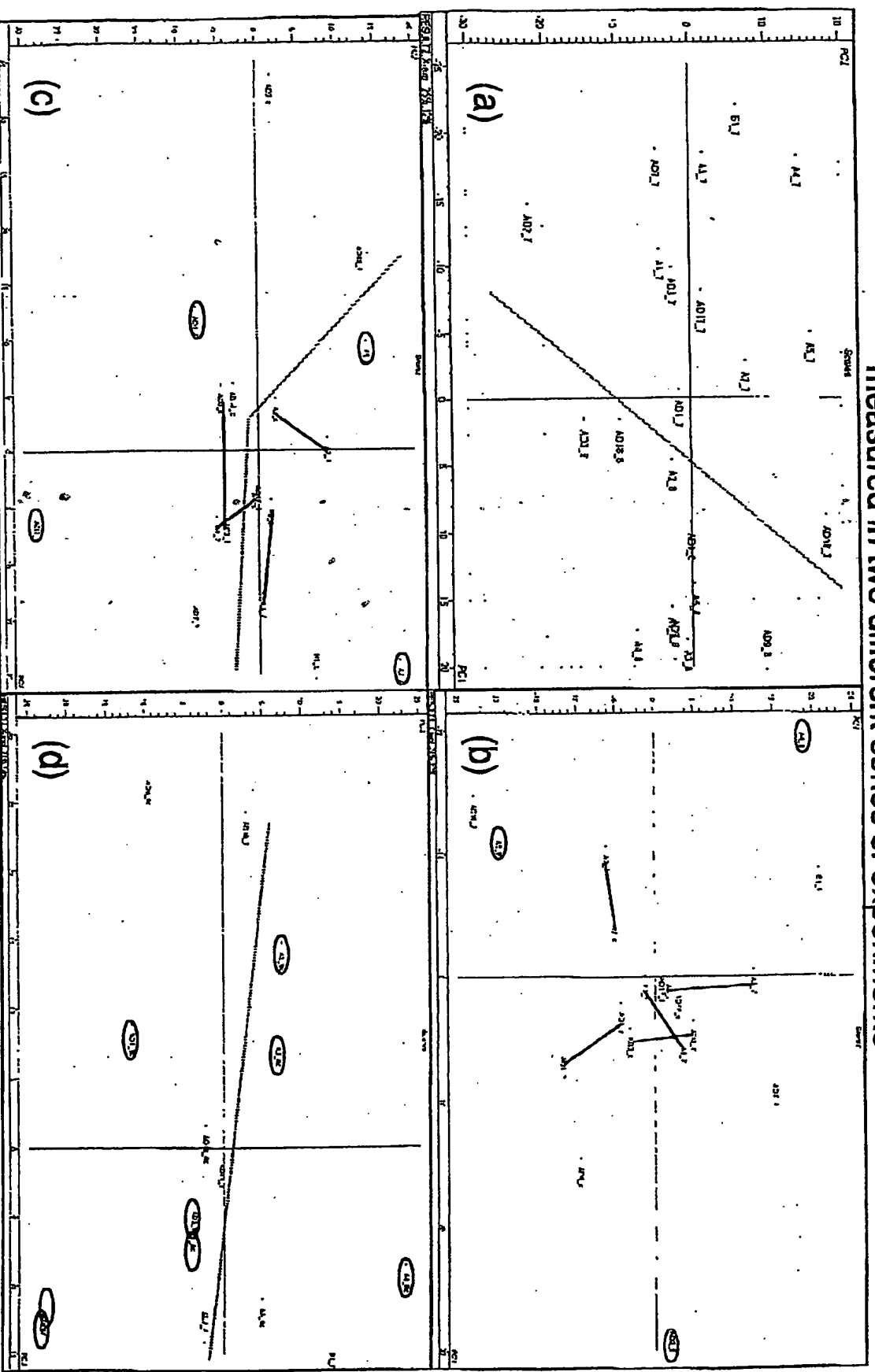


Figure 1

Effect of Direct Standardization (DS) on the Alzheimer Data measured in two different series of experiments



Description: AD (Alzheimer samples) / A.B (Non-Alzheimer samples). The samples in both series have been labeled systematically as (xx_7 / xx_8), whereas the corrected samples from series 6 (in b,c,d) have been labeled as (xx_c), thus, for example, AD2_7 denotes Alzheimer disease sample number 2 in experiment series 7. The circled spots represent the samples chosen as the transfer samples. The connecting lines in figures b,c,d show the proximity of the replicated samples after applying DS. The dashed lines in figures a,c,d represent the decision boundary separating the classes. These lines have not been drawn on the basis of any statistical criteria, but serve the purpose of visually separating the classes. All the four figures show scores plot (PC1 - PC2) from PCA analysis based on (a) Non-standardized data, (b) Scores plot after direct standardization using 3 transfer samples, (c) Scores plot after direct standardization using 4 transfer samples, (d) Scores plot after direct standardization using 8 transfer samples.

Figure 2

Projection of normal (including benign) and breast cancer samples onto a classification model generated by PLSR-DA using the data of 44 informative genes

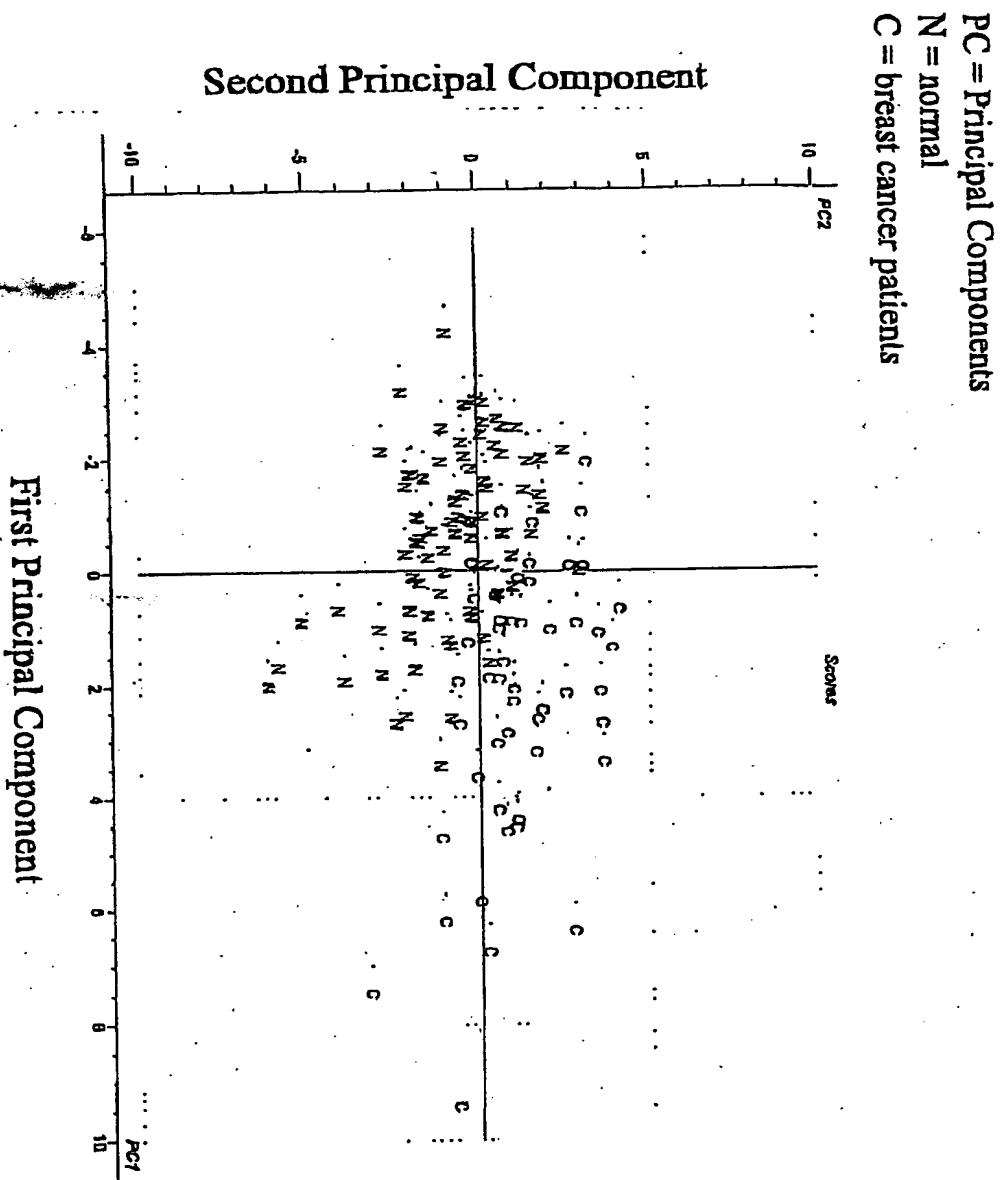
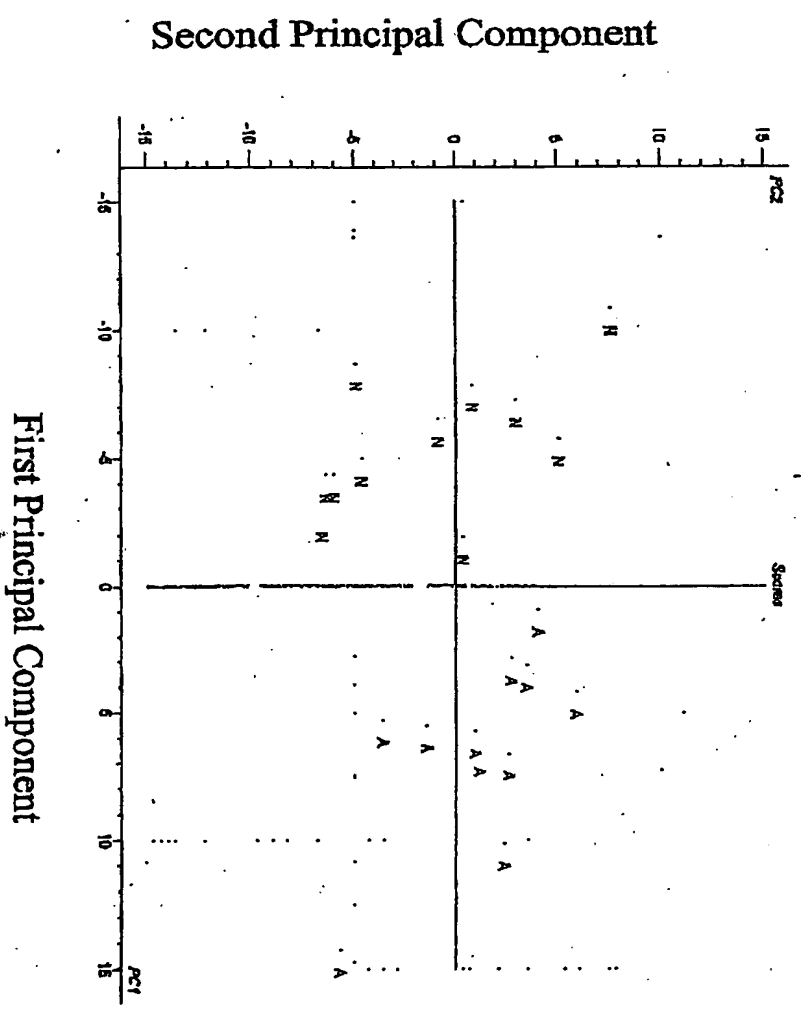


Figure 3

Projection of individuals with and without Alzheimer's disease onto a classification model generated by PLSR-DA using 182 informative genes.

PC = Principal Components
N = Non-Alzheimer's disease
A = Alzheimer's disease



PCT Application
GB0305102

